2005

Transform techniques and non-stationarity with an emphasis on network applications

Phat Piron
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Electrical and Computer Engineering Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_theses/1996

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
TRANSFORM TECHNIQUES AND NON-STATIONARITY WITH AN EMPHASIS ON NETWORK APPLICATIONS

A Thesis

Submitted to the Graduate Faculty of the Louisiana State University and Agricultural and Mechanical College in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering

in

The Department of Electrical and Computer Engineering

by

Phat Piron
M.S., University of Liège, Belgium, 2003
May 2005
“Education is an admirable thing, but it is well to remember from time to time that nothing that is worth knowing can be taught.”

Oscar Wilde
Acknowledgements

I would like to thank my major professor, Dr. Suresh Rai his help. He provided me his wise advises and pertinent remarks from time to time that have helped enhanced the contents of this thesis. I would also like to thank my committee members, Dr Subhash Kak and Dr. Morteza Naraghi-Pour. I feel honored that they gave me their consent and time to review my thesis.

I would like to express my deepest gratitude to Dung, Peng, Huey-Ying, and Yau-Si. Their constant support in all matter and at all times was invaluable. Without their help, I would not have been able to accomplish this work. Actually, without them, I would not have come to Baton Rouge. I feel blessed to have such a fantastic family. I would like to thank all other members of my family who took care of me so well.

My acknowledgements would not be complete if I would not thank my friends from all around the world. They taught me different cultures, offered me a lot of fun, and also stood with me when it was needed.

Last but not least, I would like to thank Izabela. No words would be enough to say how grateful I am to know her. Her smile and sole presence were my best motivations towards the completion of this thesis.
Table of Contents

Dedication ................................................................. ii
Acknowledgements ......................................................... iii
List of Tables ............................................................. vi
List of Figures ............................................................ vii
Abstract ........................................................................ x

1 Introduction ................................................................. 1
  1.1 Preliminaries ............................................................. 3
    1.1.1 Stationarity and Non-Stationarity .......................... 4
    1.1.2 Long-Range Dependence .................................. 6
    1.1.3 Self-Similarity ................................................ 7
  1.2 Origin of Self-similarity in Traffic ............................... 9
  1.3 Objective and Layout of the Thesis ............................. 11

2 LRD Gaussian Processes and Wavelets .............................. 13
  2.1 Fractional Brownian Motion ..................................... 13
  2.2 Fractional Gaussian Noise ....................................... 17
  2.3 Orthogonal Wavelets .............................................. 19
  2.4 Conclusion ......................................................... 23

3 Simulation and Estimation Methods .................................. 24
  3.1 Simulation Methods ............................................... 24
    3.1.1 Hosking Method ............................................ 25
    3.1.2 Davies and Harte Method ................................ 27
    3.1.3 Wavelets-Based Synthesis ............................... 30
  3.2 Estimation Methods ................................................ 33
    3.2.1 Aggregated Variance Method ............................. 33
    3.2.2 Rescaled Adjusted Range Method (R/S) ............... 35
    3.2.3 Abry-Veitch Estimator: Wavelets Method ............ 39
List of Tables

4.1 Estimate of Hurst Parameter for the Davies & Harte and the Wavelets method. 50

4.2 Abry-Veitch Estimates for Level Shift Trend. 83

5.1 Real Data: Abry-Veitch Estimator 92
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Fractional Brownian Motion for $H = 0.2, 0.5,$ and 0.8, Respectively</td>
<td>16</td>
</tr>
<tr>
<td>2.2</td>
<td>Fractional Gaussian Noise for $H = 0.2, 0.5,$ and 0.8, Respectively</td>
<td>18</td>
</tr>
<tr>
<td>3.1</td>
<td>The Aggregated Variance Method</td>
<td>35</td>
</tr>
<tr>
<td>3.2</td>
<td>The Rescaled Adjusted Range (R/S) Method</td>
<td>38</td>
</tr>
<tr>
<td>3.3</td>
<td>The Abry-Veitch Estimator: Wavelets Method</td>
<td>41</td>
</tr>
<tr>
<td>4.1</td>
<td>Performance of the Estimators for Different Values of the Hurst Parameter</td>
<td>45</td>
</tr>
<tr>
<td>4.2</td>
<td>Performance of the Wavelets Synthesis of FGN</td>
<td>49</td>
</tr>
<tr>
<td>4.3</td>
<td>Analyzed Fractional Gaussian Noise and Linear Trend</td>
<td>52</td>
</tr>
<tr>
<td>4.4</td>
<td>Performance of Estimators for a Linear Trend of Small Amplitude</td>
<td>53</td>
</tr>
<tr>
<td>4.5</td>
<td>Analyzed Fractional Gaussian Noise: Parabolic Trend</td>
<td>54</td>
</tr>
<tr>
<td>4.6</td>
<td>Performance of Estimators for a Parabolic Trend of Large Amplitude</td>
<td>55</td>
</tr>
<tr>
<td>4.7</td>
<td>Wavelets Estimator for Varying Vanishing Moments: Parabolic Trend</td>
<td>56</td>
</tr>
<tr>
<td>4.8</td>
<td>Analyzed Fractional Gaussian Noise and Polynomial Trend</td>
<td>58</td>
</tr>
<tr>
<td>4.9</td>
<td>Performance of Estimators for a Polynomial Trend of Large Amplitude</td>
<td>59</td>
</tr>
<tr>
<td>4.10</td>
<td>Wavelets Estimator for Varying Vanishing Moments: Large Amplitude</td>
<td>61</td>
</tr>
</tbody>
</table>
4.11 Performance of Estimators for a Polynomial Trend of Moderate Amplitude
4.12 Wavelets Estimator for Varying Vanishing Moments: Moderate Polynomial
4.13 Analyzed Fractional Gaussian Noise: Increasing Power-Law Trend
4.14 Performance of Estimators for an Increasing Power-Law Trend
4.15 Wavelets Estimator for Varying Vanishing Moments: Increasing Power-Law
4.16 Analyzed Fractional Gaussian Noise and Decreasing Power-Law Trend
4.17 Performance of Estimators for a Moderate Decreasing Power-Law Trend
4.18 Performance of Estimators for a Huge Decreasing Power-Law Trend
4.19 Wavelets Estimator for Varying Vanishing Moments: Decreasing Law
4.20 Analyzed Fractional Gaussian Noise: Sine Trend
4.21 Performance of Estimators for a 4-Sine Trend
4.22 Performance of Estimators for a 16-Sine Trend
4.23 Wavelets Estimator for Varying Vanishing Moments: 16-Sine Trend
4.24 Analyzed Fractional Gaussian Noise: Level Shift
4.25 R/S Analysis for Level Shift
4.26 Wavelets Estimator: 3-Level Shift
4.27 Wavelets Estimator: 5-Level Shift
4.28 Wavelets Estimator: 10-Level Shift
5.1 Data Collection Scheme
5.2 Traffic Trace
5.3 Real Data: Aggregated Variance Estimation
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4</td>
<td>Real Data: Rescaled Range Analysis</td>
<td>92</td>
</tr>
<tr>
<td>5.5</td>
<td>Traffic Trace: Zoom on Beginning</td>
<td>94</td>
</tr>
<tr>
<td>5.6</td>
<td>Traffic Trace: Zoom on Middle</td>
<td>95</td>
</tr>
<tr>
<td>5.7</td>
<td>Traffic Trace: Zoom on End</td>
<td>96</td>
</tr>
</tbody>
</table>
Abstract

The recent years brought a phenomenal development of Internet. It is, therefore, important to find some ways to improve the performances. The first step in this direction is the characterization and modeling of the network traffic. It has been tested that the network traffic behaves like a self-similar process, while packets interarrivals time possess the long-range dependence property. In particular, we model them by using fractional Brownian motion and fractional Gaussian noise, respectively. Note that, the former is just the cumulative sum of the latter. By using these concepts, the traffic characterization reduces to the estimation of one value: the Hurst parameter. Numerous methods exist to evaluate this parameter. Nevertheless, a few studies take account of the inherent non-stationarity present in real data. For short samples, the stationarity hypothesis might hold. But for larger samples, this is hardly the case. As an example, for network traffic, the day cycle shows non-stationarity. By not considering the non-stationarity, an inaccurate or even inappropriate estimation may result. Our objective in this thesis is to test the robustness of several techniques such as aggregated variance method, rescaled range method, and wavelets method, in presence of a set of non-stationarity trends. We study the estimators on a known signal generated using Hosking, Davies and Harte method, or wavelets-based synthesis. We
add various deterministic non-stationarity trends to the original signal. We considered polynomial, power-law, sinusoidal, and level-shift trends. Results help analyze the behavior of the estimators. All the simulations are carried out using Matlab\textsuperscript{1}. We show that, depending on the trend, the estimators react differently. We have also used real data to verify the effectiveness of estimators. Results confirm the observations that we have made with lab data. In particular, we show that the wavelets method provides several flaws. Especially, its results must be carefully analyzed when the data is non-stationary.

\textsuperscript{1}Matlab is a trademark of The Mathworks, Inc
Chapter 1

Introduction

Internet applications has become more and more common in every day life. The past few years have seen an exponential increase in both the number of users and the number of services offered to users. Applications require more and more resources, while the users are expecting a certain degree of quality. A key problem for Internet is, therefore, to identify some ways to improve or at least maintain the performance of network communications. The first step towards a better understanding of how to analyze and control the performance of a network is the characterization and modeling of its traffic.

For a long time, packets arrival were approximated by a Poisson process, because they offered a nice set of statistical properties. Nevertheless, it has been shown that packets interarrivals distribution is not exponential and utilization of this model is incorrect [22] [21]. The correlation function of a Poisson process decays exponentially. It means that a Poisson process has a short-range dependence (SRD). In other words, there is no correlation between
events separated by a sufficient period of time. As a consequence, the bursty behavior of the traffic (i.e. the traffic sharp fluctuation on different time scales) would be able to repeat only on a small time interval. This explains the most important shortcoming of the Poisson modeling, which is not able to handle efficiently the bursty behavior of Internet traffic.

While Poisson processes have the SRD property, measurements of networks traffic over large scales provide a better understanding of the traffic behavior. References [8, 22, 24] show that packets arrivals are not independent over a large time scale. In other words, packets arrivals exhibit long-range dependence (LRD). This means that when viewed on a sufficiently large time scale, the cumulative traffic trace can be interpreted as a self-similar process. Simply put, self-similarity means that if we compare a process at a certain time scale and the same process at another time scale, the latter looks like a scaled version of the former. Self-similarity provides a nice framework for the characterization and modeling of the cumulative traffic. For example, it can handle easily the extended range of burstiness inherent in network traffic. By using the notions of self-similarity and long-range dependence, the traffic can be modeled with the so-called Hurst parameter. Details on self-similarity are more thoroughly explained in Section 1.1.3.

In order to be able to characterize network traffic correctly, one need to estimate the Hurst parameter as accurately as possible. Several methods are available to perform this task with various precisions, performances, and limitations. As a result, a lot of studies has been conducted to find evidence of the self-similar nature of Internet traffic. However, most studies are carried out under the stationary process assumption. This is an important hypothesis
since the estimators utilize this notion to evaluate the value of the Hurst parameter. In reality, stationarity hold on on short time period. It is hardly the case on larger time scale. For example, it is observed that the day traffic pattern is different from the night traffic pattern. Some differences also exist depending on the hour, day, and month. This non-stationarity can easily compromise the estimator and introduce a bias in the value of the Hurst parameter. At worse, the estimator could conclude to long-range dependence where there is none.

This chapter defines more clearly the concept of long-range dependence and self-similarity. It also discusses the origins and causes of self-similarity in the traffic. Finally, we outline the aim and layout of this thesis.

1.1 Preliminaries

This section provides basic notion and some mathematical background about various terms such as non-stationarity, long-range dependence, and self-similarity.

Consider a discrete-time process \( \{X_t, t \geq 0\} \). We define the following term as:

- The mean of the process is
  \[
  m(X_t) = \mu_{X_t} = E[X_t]
  \]  \hspace{1cm} (1.1)

- The variance of \( X_t \) is
  \[
  Var(X_t) = \sigma^2_{X_t} = E[(X_t - \mu_{X_t})^2].
  \]  \hspace{1cm} (1.2)

- The autocovariance of \( X_t \) is
  \[
  C_X(t, s) = \gamma_X(t, s) = E[(X_t - \mu_{X_t})(X_s - \mu_{X_s})].
  \]  \hspace{1cm} (1.3)
Finally, the autocorrelation is

$$R_X(t, s) = \rho_X(t, s) = \frac{\gamma_X(t, s)}{\sigma_X(t) \sigma_X(s)}.$$  \hspace{1cm} (1.4)

Note that the second definition refers to the variance of a process. It should then be referred as the autovariance. This study always use the variance in the context of a process. For the remaining sections variance and autovariance both refer to the same value, unless stated otherwise.

We see that the autocorrelation and autocovariance functions have simpler definitions as they depend only on the time difference $k = t - s$.

1.1.1 Stationarity and Non-Stationarity

We define a discrete time process $\{X_t, t \geq 0\}$ as stationary if the extracted samples $X_{t_1}, \ldots, X_{t_n}$ and $X_{t_1+s}, \ldots, X_{t_n+s}$ have the same distribution for every $n, s, t_1, \ldots, t_n$. It means that two samples of the same length have the same distribution, independently of the origin of the samples. This is why such processes are also called shift-invariant processes [25, 26].

This definition is pretty strict and limits the number of processes to include in this category. We use a less stringent definition of stationarity called second-order stationarity or weak stationarity. It requires that the first two moments and the autocovariance do not depend on time. More precisely mean and variance are constant, i.e. $\mu_{X_t} = \mu_X$ and $Var(X_t) = Var(X)$. It implies that the autocovariance between samples $X_t$ and $X_s$ depend only on the value of $|t - s|$. Therefore, if $k = |t - s|$ we can now write the autocovariance as $\gamma_X(t, s) = \gamma_X(k)$. In our study, we concentrate essentially on second-order non-stationary
processes. Such a process is entirely defined by its mean and variance (or autovariance). Gaussian processes obviously belong to this category.

The advantage of the stationary processes is obviously the simplicity of the expression of their moments. If \( k = |t - s| \) and the autocovariance is described by \( \gamma_X(k) \), the autocorrelation of equation (1.5) is rewritten as

\[
\rho_X(k) = \frac{\gamma_X(k)}{\sigma^2_X}. 
\]  

(1.5)

For a process to be non-stationary, it suffices for one of its moments to depend on time. On large time scale, it is difficult to believe that the stationarity hypothesis hold. As an obvious network traffic example, we can take the diurnal cycles and the peak hours, which are modifying significantly the behavior of the traffic. The problem is that, strictly speaking, there is an infinite number of possibilities to describe non-stationarity. Thus, it is difficult to define a proper way to analyze a signal under such conditions. This is why we shall limit in this thesis the non-stationarity to a popular set called trends.

We consider a stationary process to which another signal is added. Our original process might then look like a non-stationary one. This parasite signal can have any form and may threaten any analysis we would like to perform on the original signal. In that case, the analyzed signal \( X_t \) can be decomposed in two parts, the original signal \( S_t \) and the perturbing trend \( P_t \). Mathematically, we have

\[
X_t = S_t + P_t. 
\]  

(1.6)

The trend \( P_t \) can take many different forms. It can be a sudden change of mean, i.e. a level shift. It can also be a linear or parabolic function, or even any polynomial functions of higher
degree. It can also be an increasing or decreasing power-law \((t^\beta)\). Besides all these, we also consider a trigonometric function where \(P_t\) is a sine function. Note, the perturbing trend can take many other forms, but we believe that those functions represent a good set of different behaviors. One of the reasons behind the choice of trends to model the non-stationarity is that it is a good approximation of real cases. For example, a peak hour or a Denial of Services (DOS) attack on a network can often be seen as transient level shifts. Also, the diurnal cycle can be approximated as a polynomial trend.

1.1.2 Long-Range Dependence

As discussed earlier, the short-range dependence refers to a process with an autocovariance function that decays exponentially fast. This, in turn, leads to uncorrelatedness between events occurring in sufficiently spaced time period. Long-range dependence occurs when an autocorrelation function decays more slowly, following a power law. A long-range dependence process has \(\gamma_X(k) \sim c_r|k|^{-(1-\alpha)}, k \to \infty, \alpha \in (0, 1)\), \(1.7\)

where \(\alpha\) is a dimensionless scaling exponent and \(c_r\) denotes a quantitative parameter that gives a measure of the magnitude of LRD induced effects. From \(1.7\), it follows that \(\sum_{k=0}^{\infty} \gamma_X(k) = \infty\). This property contrasts to short-range dependence processes, which possess finite summable autocovariances. Note that the definition of LRD is only an asymptotic definition that takes into account the behavior at infinity. Therefore, even though individual autocorrelations can be arbitrarily small, their sum is always significant. It means the past has some influence on the future of the process. This leads to a higher variability, a
main issue for the proper statistical analysis and estimation. In (1.7), we use \( c_r \), but in the literature \( c_f \) is sometimes used. It does not make much difference, as these two parameters are linked by the following relation [3]:

\[
    c_f = 2(2\pi)^{-\alpha} c_r \Gamma(\alpha) \sin((1 - \alpha)\pi/2),
\]

where \( \Gamma \) refers to the Gamma function.

An illustration of the influence of the LRD parameters is given in [27]. Assume we have a stationary process \( X_t \) of length \( n \). Generally, we have a sample mean with a uniform distribution with expectation \( \mu_X \) and variance \( Var(X)/n \) for large \( n \). But in the case of long-range dependence, we find the variance to be \( \frac{2c_r n^\alpha}{(1+\alpha)n} \). We immediately notice the result depends on \( \alpha \) and \( c_r \). This result also implies slower rate of decrease with \( n \) for LRD. This means the confidence intervals for large \( n \) are better than that in usual cases.

Further we define long-range dependence using the autocorrelation as [27]

\[
    \rho_X(k) \sim c_r |k|^{-(1-\alpha)}, k \to \infty, \alpha \in (0, 1),
\]

where \( c_r \equiv c_r/Var(X) \) is a dimensionless constant.

1.1.3 Self-Similarity

Self-similar processes were first introduced by Kolmogorov in 1941 [21]. Nevertheless, they did not have much impact on statistics until Mandelebrot re-introduced and developed them during mid-seventies [23] in fractal and multi-fractal studies.

As the name suggests, a self-similar process looks or behaves alike when viewed at different time scales. More accurately, say we have a stationary process \( Y_t \) and an arbitrary positive factor \( m \). \( Y_t \) is self-similar with Hurst parameter \( H \), if \( Y_t \) and its rescaled version at time
scale \( mt \), say \( m^{-H}Y_{mt} \), have the same distribution. Mathematically [13],

\[
\{Y_t, \; t \in \mathbb{N}\} =_{d} \{m^{-H}Y_{mt}, \; t \in \mathbb{N}\}, \; \forall m > 0,
\]

where \( H \in (0, 1) \) and \( =_{d} \) denotes equivalence in distribution. An important property of the self-similar process is related to the scaling of their moments [30]:

\[
E|Y_t|^q = E|Y_1|^q t^{qH}.
\]

Equation (1.10) relates the \( q^{th} \) moment, if it exists, to a power-law function. No matter which time scale we observe, the process has the same fluctuations. Since the moments are all dependent on time, it explains why self-similar processes are not stationary.

Reference [3], the author notes another important characteristic of self-similar processes. A key concept of self-similar processes is that they have stationary increments. By stationary increments we mean that the number of events occurring at a given period depends only on the length of this period. As a particular case, we may sum a long-range dependent stationary stochastic process to obtain the self-similar process. In this case, it can be shown that \( \alpha \) and \( H \) are related following \( H = (1 + \alpha)/2 \). Because of the strong relation between those two types of processes, long-range dependence is sometimes also determined in terms of \( H \) rather than in term of \( \alpha \). In fact, we can rewrite equation (1.7), which defines long-range dependence, as [27]

\[
\gamma_X(k) \sim c \times k^{2(H-1)}, \; k \to \infty, \; H \in \left( \frac{1}{2}, 1 \right).
\]

Finally, it is important to remark that even though long-range dependence and self-similarity are strongly related, they are not equivalent. For example, it is possible to have a
self-similar process that does not possess increments having a long-range dependence. An ob-
vious and famous process with this property is the classical Brownian motion. The Brownian
motion is self-similar with Hurst parameter as $1/2$, while its increments are independent.

1.2 Origin of Self-similarity in Traffic

Many studies have proved that network traffic is self-similar [8, 22, 24]. While we are trying
to analyze the traffic from this perspective, one may wonder about the origin of self-similarity
in network traffic. This is a very wide subject and is still under research.

One interesting paper treating this subject is [8], in which the authors analyze traffic from
the World Wide Web (WWW). They note that to perform a correct and complete analysis,
they have to consider every aspects from the traffic behavior to the users decisions. They call
this the genesis of Web traffic. In order to complete the analysis, they try to decompose the
traffic by multiplexing ON/OFF sources that are heavy-tailed. A distribution is heavy-tailed
if its distribution follows power-law as

$$P[X > x] \sim x^{-a}, \text{ as } x \to \infty, \ 0 < a < 2. \quad (1.12)$$

From their study, it is shown that such sources lead to self-similarity. For the ON periods,
the heavy-tail comes from the distribution of the available file sizes, while for the OFF
periods this is due to the user “think time.” They conclude that the self-similarity is not a
“technological consequence.” This property should therefore pretty much hold if we modify
the medium of transport, for instance, by changing TCP.

Acknowledging the origins of the self-similarity, some studies try to discern the exact role
of TCP in self-similarity. For example, [15] [16] show that TCP has an important influence
on short time scales, but does not have any effects on large time scales. Authors in [32] do not share the opinion. They argue that this later assumption is correct only if we treat the traffic locally on a single link. On the contrary, they show that by considering the whole network the conclusion is different. They also find that the end-to-end congestion control of TCP is sensitive to any time scale, which means that TCP actually have an influence on large time scale behavior. In particular, they conclude that TCP propagates the long-range autocorrelation of the traffic through the bottlenecks. The largeness of the time scale depends on many parameters as round-trip time, window size, and speed of connection.

In [14], authors emphasize to understand why studies about TCP have different conclusions. Their explanation holds to the fact that the analysis is biased from the start. In effect, TCP is a closed-loop algorithm. It means that some information at the end of the communication is transmitted and affects the source. However, most of the studies are considering TCP as an open-loop process. This signifies that the performance at the end of the relation is not modeled. In order to prevent this, they propose an approach based on chaotic maps. A chaotic map is the simplest form of chaos. Chaos is characterized by three simple ideas. It is deterministic, sensitive to initial condition, and has underlying patterns sometimes called attractors. In other word, it means that the short term behavior may be predicted, but not the long term one, with the attractors having an important influence on both behavior. Also, chaos is generally thought of as low-dimensional, meaning it can be described by a few simple variables. We encourage the interested readers to consult [14]. Their conclusion is that although TCP, and in particular its feedback, can modify the self-similarity of the
traffic, it can neither generate or destroy it.

1.3 Objective and Layout of the Thesis

The objective of this thesis is to see how we can analyze traffic traces in the presence of non-stationarity. We study three different techniques that estimate the Hurst parameter. We test their performances under some specific non-stationary conditions. In particular, we test the results obtained by the wavelets technique, considered by many papers as the most accurate and the most robust. Results show that this method has several flaws. We notice that the estimation of a non-stationarity polynomial trend is correct only in a limited number of cases. We also observe that wavelets technique is not robust to one of the most important type of non-stationarity, the level-shift. Further, we found that another method, not considered as good as the wavelets method, actually perform well under non-stationary conditions. Finally, we verify our conclusions from the simulations using real data from the network obtained from *tcpdump*.

The layout of the thesis is as follows. Chapter 2 presents Gaussian processes with long-range dependence. We assume that the traffic is self-similar and that its packets interarrivals times have long-range dependence. This chapter also provides a discussion on orthogonal wavelets, which are useful in the study of self-similarity and long-range dependence. In Chapter 3, we consider various tools needed to estimate the Hurst parameter ($H$). First, we describe three methods to generate the self-similar process for a given $H$. These methods provide data in our analysis in Chapter 4. Second, we present three estimators that are used to evaluate $H$. All techniques are implemented in Matlab. Chapter 4 applies concepts
presented in Chapter 3. Here, we verify the exactness of the estimators with no interfering signals. We, next, add different type of trends (such as polynomial, power-law, sinusoidal, or level shift) to the signal to simulate non-stationarity. We observe the performances of the estimators in the presence of non-stationarity. Chapter 5 analyzes data from a real network obtained using tcpdump. Results obtained in Chapter 4 confirm our conclusion in the real data case and confirm our previous conclusions. Finally, Chapter 6 concludes the thesis and provides directions for the future work.
Chapter 2

LRD Gaussian Processes and Wavelets

This chapter discusses Gaussian type long-range processes and studies their typical properties. It starts with a fractional Brownian motion process, which is self-similar with Hurst parameter $H = 1/2$. It, then, describes fractional Brownian motion process and fractional Gaussian noise process. These processes are used/referenced quite frequently in the remaining of the thesis.

To complete our discussion, we have also given key concepts of orthogonal wavelets. This includes basic ideas of multiresolution analysis, regularity, and vanishing moments.

2.1 Fractional Brownian Motion

We first define a Brownian motion process. A stochastic process $\{B(t), t \geq 0\}$ is said to be a Brownian motion process if \cite{26}:

1. $B(0) = 0$

2. $\{B(t), t \geq 0\}$ has stationary and independent increments and
3. for every $t > 0$, $B(t)$ is normally distributed with mean 0 and variance $\sigma^2 t$.

From these, we conclude that both the increments and the Brownian motion are Gaussian. Mathematically, properties 2 and 3 can be expressed as

\[ E[B(t) - B(s)] = 0 \]  \hspace{1cm} (2.1)

\[ Var[B(t) - B(s)] = \sigma^2 |t - s| \]  \hspace{1cm} (2.2)

Using (2.1) and (2.2), it is simple to verify self-similarity. Note

\[ E[B(ct)] = E[B(ct) - B(0)] = 0 = c^{\frac{1}{2}} E[B(t)]. \]

The increments are independent and

\[ \gamma_B(t, s) = Var(B(s) - B(0)) = \sigma^2 s, \]

if $t > s$. Then for any $c > 0$

\[ \gamma_B(ct, cs) = c \sigma^2 s = \gamma_B(c^{\frac{1}{2}} B(t), c^{\frac{1}{2}} B(s)). \]

Since we have a Gaussian process, we can conclude that $B(t)$ is self-similar, with $H = \frac{1}{2}$.

Now, consider a self-similar process $Y_t$. It follows $Y_t$ has stationary increments. The mean of increments $X_t = Y_t - Y_{t-1}$ is zero. The autocovariance is [5]:

\[ \gamma_X(k) = \frac{1}{2} \sigma^2 [(k + 1)^{2H} - 2 k^{2H} + (k - 1)^{2H}] \]  \hspace{1cm} (2.3)

where $\sigma^2 = Var(X)$. Moreover, increments are Gaussian. In other words, it is a second-order process. Therefore, it is totally determined since we know its mean and the variance.
It means for each value of $H$ we find one Gaussian process $X_t$ related to the self-similar process $Y_t$. In such a case, the increments $X_t$ are called fractional Gaussian noise, and their cumulative sum $Y_t$ is called fractional Brownian motion. We denote the latter as $B_H(t)$. As we stated earlier, if $H = 1/2$, $B_H(t)$ is a simple Brownian motion denoted by $B(t)$.

Fractional Brownian motion process is defined by summing fractional Gaussian noise, which have autocovariance given in (2.3). Another way to define $B_H(t)$ is by computing a weighted average of $B(t)$ over infinite past using [13]

$$B_H(t) = C_H \left\{ \int_{-\infty}^{0} [(t - s)^{H-1/2} - (-s)^{H-1/2}] dB(s) + \int_{0}^{t} (t - s)^{H-1/2} dB(s) \right\}, \quad (2.4)$$

where

$$C_H = E[B_H(1)^2]^{1/2} \left\{ \int_{-\infty}^{0} ((t - s)^{H-1/2} - (-u)^{H-1/2})^2 du + \frac{1}{2H} \right\}^{-1/2}.$$ 

Note that if we choose $H = 1/2$, the increments are independent and the result is the simple Brownian motion. The reader interested in how to solve this equation and in how to derive the properties may refer to [13]. The main result is that the variance of $B_H(t)$ is $V_H t^{2H}$.

In particular, we define a normalized fractional Brownian motion process having following properties

1. $B_H(t)$ has stationary increments
2. $B_H(0) = 0$ and $EB_H(t) = 0$ for $t \geq 0$
3. $EB_H^2(t) = t^{2H}$ for $t \geq 0$
4. $B_H(t)$ has a Gaussian distribution for $t > 0$. 

15
From the first three properties, we find that the autocovariance function is given by \[5\]

\[\gamma(s,t) = EB_H(s)B_H(t) = \frac{1}{2}t^{2H} + s^{2H} - (t-s)^{2H}\]

for \(t > s\). Since this is a second-order process, it is totally determined from its first two moments. It is then fairly easy to verify that \(B_H(t)\) is self-similar. Notice the fractional Brownian motion process is the only Gaussian process with stationary increments that is self-similar.
Realizations of fractional Brownian motion are illustrated in Figure 2.1. We simulated them for Hurst parameter of \{0.2, 0.5, and 0.8\}. Observe the fluctuations for \(H = 0.2\) and for \(H = 0.8\). In former they are noticeablem while in latter they are not. To explain this difference, it is easier to introduce first the increment process, i.e. the fractional Gaussian noise. It is also used to build the fractional Brownian motion.

2.2 Fractional Gaussian Noise

We already defined the fractional Gaussian noise, \(X_k\), as the stationary increment of the fractional Brownian motion, \(B_H(t)\). Mathematically,

\[
X_k = B_H(k + 1) - B_H(k).
\]

(2.6)

Note, has a standard normal distribution. It is independent only when the autocovariance function given in (2.3) is zero, i.e. when \(H = 1/2\). Their sum then yields a normal Brownian motion process. For \(H > 1/2\), \(X_t\) has long-range dependence.

From (2.3), we can also derive the actual long-range dependence by obtaining an expression similar to (1.11). We consider the Taylor expansion at the origin of the function

\[
h(x) = (1 - x)^{2H} - 2 + (1 + x)^{2H},
\]

and we note that \(\gamma(k) = \frac{1}{2}k^{2H}h(1/k)\) for \(k \geq 1\). Thus, it follows \[5\]

\[
\gamma(k) \sim H(2H - 1)k^{2H-2}
\]

(2.7)
as \(k \to \infty\). Equation (2.7) states that for \(H > 1/2\), a fractional Gaussian noise has long-range dependence.

Finally, we make some interesting observations about fractional Gaussian noise. First,
it is easy to show that fractional Gaussian noise is actually self-similar \[12\]. This is done by developing the autocovariance function and using the relation with fractional Brownian motion. Second, from (2.3) we see that the autocovariance is negative for \( H < 1/2 \). This negative autocorrelation leads to a process with a lot of fluctuations, while a process with a positive autocorrelation have some increasing period and some decreasing period. This last observation is verified in Figure 2.2 which illustrates simulations of fractional Gaussian noise for a Hurst parameter of \( \{0.2, 0.5, \text{ and } 0.8\} \) from top to bottom. The first plot oscillates between positive and negative values, while the last plot retains some periodicity. From

Figure 2.2: Fractional Gaussian Noise for \( H = 0.2, 0.5, \text{ and } 0.8, \text{ Respectively} \)
these observations, we have a better understanding of the plots in Figure 2.1. For a small Hurst parameter, the process has small autocorrelations and, therefore, there are a lot of oscillations. On the other hand, with a larger Hurst parameter, we observe autocorrelations between the samples, which leads to a smoother plot.

2.3 Orthogonal Wavelets

Basically, the idea behind wavelets transformation is similar to the one that carries the Fast Fourier Transform (FFT). Both techniques aim to decompose the signal through a particular basis, providing an approximation of the signal that highlights some characteristics. In case of FFT, where bases are trigonometric sine and cosine functions, the signal is transformed into the frequency domain. But this is done at the cost of time resolution. With wavelets analysis, we may use different sets of bases as long as they follow some properties. Wavelets allow more flexibility for the signal analysis. In fact, with wavelets, the signal is transformed into a time scaled wavelet domain, where the bases are dilatations and translations of the original mother wavelet $\psi_0$. The latter works as a band pass function which limit the signal, both in frequency and time. Therefore, we can say that, while conserving the time dimension of the original data, the wavelets allow to observe a time series at different scales. This property is particularly useful in the fractional Brownian motion analysis.

In order to describe the wavelets more formally, we need to go through some definitions. Refer to [10] for thorough concept on the topic. Let $L^2(\mathbb{R})$ be the space of all square-integrable functions on $\mathbb{R}$. In $L^2(\mathbb{R})$, functions are distributed into classes. Two functions are in the same class if they are almost everywhere equal in the sense of a countable set of
points. If \( f, g \in L^2(\mathbb{R}) \), we define the inner-product

\[
\langle f, g \rangle = \int_{\mathbb{R}} f(t)g(t)dt.
\] (2.8)

We can also define the \( L^2 \)-norm as \( \|f\|^2 = \langle f, f \rangle \). If for the set \( \{f_k, \ k \in K\} \), we have

\[
\langle f_i, f_j \rangle = 1_{i=j}, \ \forall i, j \in K,
\]

where \( 1_{i=j} = 0 \) unless \( i = j \), then the set of functions is said to be orthonormal. Finally, we define a basis for \( L^2(\mathbb{R}) \). Say we have a subset \( A \in L^2(\mathbb{R}) \) and a set of functions \( \{f_k : k \in K\} \). The latter is a basis of \( A \) if we can uniquely write

\[
f = \sum_{k \in K} \alpha_k f_k, \ \forall f \in A,
\] (2.9)

for some value \( \alpha_k \). Furthermore, note that in the case where \( K \) is a countable set, the relation is stronger. For example, if \( K = \mathbb{Z} \) then \( f = \sum_{k \in \mathbb{Z}} \alpha_k f_k \) is the unique function

\[
\lim_{N \to \infty} \|f - f = \sum_{k=-N}^{N} \alpha_k f_k \| = 0.
\]

Finally, note results hold for weaker basis, called Riesz bases.

We now define the multiresolution analysis (MRA). The multiresolution analysis is the analysis of a signal using wavelets bases. It is defined by a function \( \phi_0 \), called the scaling function, and a collection of nested subspaces \( \{V_j : j \in \mathbb{Z}\} \), with \( V_j \in L^2(\mathbb{R}) \). The subspaces have the following properties [1, 10, 12]:

1. \( \bigcap_{j \in \mathbb{Z}} V_j = 0 \);
2. \( V_j \subset V_{j-1} \) for \( j \in \mathbb{Z} \);
3. Any function \( f \in L^2(\mathbb{R}) \) can be approximated by functions in \( V_j \) to arbitrary precision in \( L^2 \)-norm, it means that there exists a basis of \( L^2(\mathbb{R}) \) in \( V_j \);
4. \( x(t) \in V_j \) if and only if \( x(2^j t) \in V_0, \forall t; \) and

5. \( \phi_0 \in V_0 \) and \( \{ \phi(t - k) : k \in \mathbb{Z} \} \) form an orthonormal basis for \( V_0 \).

Also, it can be shown that the scaled and shifted scaling functions

\[
\{ \phi_{j,k}(t) = 2^{-2j/2}\phi_0(2^{-j} t - k) : k \in \mathbb{Z} \}\]

form an orthogonal basis for \( V_j \).

A multiresolution analysis is done by successively projecting the analyzed signal, say \( x \), on the subspaces \( V_j \). Those projections are called approximations,

\[
\text{approx}_j(t) = (\text{Proj}_{V_j} x)(t) = \sum_{k \in \mathbb{Z}} a_x(j, k) \phi(j, k)(t). \tag{2.11}
\]

Because \( V_j \subset V_{j-1} \), each projection on a smaller set provides a coarser approximation of the original signal. The idea behind the multiresolution analysis is to study the loss of quality between the different approximations. This difference is called detail, such that we have \( \text{detail}_j(t) = \text{approx}_{j-1}(t) - \text{approx}_j(t) \). It can be checked that details are computed directly rather than going through the approximations. In effect, define \( W_j \) as the subspace \( V_{j-1} \) from which we remove \( V_j \), i.e. \( W_j = V_{j-1}/V_j \). Since \( V_j \) and \( W_j \) are orthogonal spaces, \( \{\forall f \in W_j, \forall g \in V_j, \langle f, g \rangle = 0 \} \), \( \text{detail}_j(t) \) can be computed by projecting the signal on \( W_j \).

It can be checked that there exists a function \( \psi_0 \), derived from \( \phi_0 \), such that

\[
\{ \psi_{j,k}(t) = 2^{-j/2}\psi_0(2^{-j} t - k) : k \in \mathbb{Z} \}\]

is an orthonormal basis for \( W_j \). It follows

\[
\text{details}_j(t) = (\text{Proj}_{W_j} x)(t) = \sum_{k \in \mathbb{Z}} d_x(j, k) \psi(j, k)(t). \tag{2.13}
\]
Therefore, through multiresolution analysis, we can rewrite the signal with a low-resolution version of the signals and the set of details that denote loss from the original signal $x(t)$:

$$x(t) = \text{approx}_J(t) + \sum_{j=1}^{j=J} \text{detail}_j(t).$$  \hspace{1cm} (2.14)

By inserting equation (2.11) and (2.13) in equation (2.14), we obtain the final form of the wavelet decomposition at level $J$:

$$x(t) = \sum_{k \in \mathbb{Z}} a_x(J,k)\phi(J,k)(t) + \sum_{j=1}^{j=J} \sum_{k \in \mathbb{Z}} d_x(j,k)\psi(j,k)(t).$$  \hspace{1cm} (2.15)

Since the approximations lead to less and less details, $\phi_0$ is seen as low-pass filter. On the other hand, the “differential” information contained in the details indicate that $\psi_0$ is a bandpass function or small wave, or wavelet. Note that the coefficients of equation (2.15) are computed through the inner product of $x$ with, respectively, the scaling and the mother function:

$$a_x(j,k) = \langle x, \phi_{j,k} \rangle$$  \hspace{1cm} (2.16)

$$d_x(j,k) = \langle x, \psi_{j,k} \rangle$$

We point out another advantage of wavelets. Their decomposition can be computed in a fast, efficient, and low-cost fashion because of a recursive filter-bank pyramid algorithm [10].

Further examinations show that, because of regularity issues, $\psi_0$ should satisfy $\int \psi_0(t)dt = 0$. Many of the wavelets have [1]

$$N(k) = \int t^k \psi_0(t)dt = 0 \text{ for } k = 0, 1, ..., N - 1,$$  \hspace{1cm} (2.17)

where $N$ is a positive integer called the number of vanishing moments for the wavelet function. This is one of the most important features of the wavelets. More accurately, the
regularity of $\psi(t)$ is defined in [33]. If $\psi(t)$ is $N$-times differentiable and decays fast enough, then the first $(N-1)$ wavelet moments vanish. That is

$$\left| \frac{d^k\psi(t)}{dt^k} \right| < \infty; \quad 0 \leq k \leq (N-1)$$

implies $N(k) = 0$ for $0 \leq k \leq (N-1)$. Unfortunately, the converse is not true. For example, the $DAUB20$ [10] is 3 times differentiable. Following (2.18), it has at least 3 vanishing moments. Actually, it has 10 vanishing moments. Another example is from $DAUB6$. It has 3 vanishing moments, but it possesses one time differentiable. In Chapter 4, we see that this relationship between regularity of $\psi(t)$ and the number of vanishing moments have important consequences.

2.4 Conclusion

This chapter considers Gaussian type LRD processes. Specifically, we described the fractional Brownian motion to model the traffic and the fractional Gaussian noise to model the packets interarrivals time. We have also described basic orthogonal wavelets. In this, both decomposition and reconstruction require only one set of filter coefficients. Vanishing moments of wavelets function relate to the differentiability of a function and help detect singularities.
Chapter 3

Simulation and Estimation Methods

To characterize the signal, we need to find the value of its Hurst parameter. A number of methods are available to perform this estimation with different advantages and shortcomings. In order to discern the features of the estimators, it is good to first test them on data where $H$ is known. In Section 3.1 we describe a few techniques to simulate fractional Gaussian noise, which we can use, in turn, to create fractional Brownian motion. Finally, Section 3.2 present the estimators chosen to evaluate the Hurst parameter. Note that all the procedures described in this chapter have been programmed and tested using Matlab.

3.1 Simulation Methods

It must be noted that this section has received a very important contribution from the work compiled in [12]. We review (a) Hosking method, (b) Davies and Harte method, and (c) wavelets-based synthesis method. We implemented the first two methods in Matlab. The wavelets-based method is available with Matlab 7.0. Also note that all the methods presented
here are defined in discrete time.

3.1.1 Hosking Method

Sometimes called Durbin or Levison, the Hosking method is an algorithm to compute fractional Gaussian noise [18]. From there, we just have to compute the cumulative sum of the result to obtain the fractional Brownian motion. This method is based on the key fact that a particular sample can be completely computed given its past. In other words, given $X_0, \ldots, X_n$, we are able to determine $X_{n+1}$. From this result, we then just have to apply the algorithm recursively until we produce enough samples.

The autocovariance for FGN is defined in equation (1.3). In this particular case, we use discrete time and analyze a Gaussian process. If we call our process $X$, the mean of $X$ is zero. We can then simplify the autocovariance to:

$$\gamma_X(k) = \gamma(k) = EX_nX_{n+k}, \text{ for } n, k = 1, 2, \ldots$$  \hspace{1cm} (3.1)

Use $\gamma(0) = 1$. We define (a) the covariance matrix as $\Gamma(n) = (\gamma(i-j))_{i,j=0,\ldots,n}$, (b) the column-vector $c(n)$ where $c(n)_k = \gamma(k+1)$ for $k = 0, \ldots, n$, and, (c), the matrix $F(n) = (1(i = n - j))_{i,j=0,\ldots,n}$ where $1$ denotes the indicator function. Thus

$$\Gamma(n) = \begin{pmatrix}
\gamma(0) = 1 & \gamma(1) & \gamma(2) & \cdots & \gamma(n) \\
\gamma(1) & 1 & \gamma(1) & \cdots & \gamma(n-1) \\
\gamma(2) & \gamma(1) & 1 & \cdots & \gamma(n-2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma(n) & \gamma(n-1) & \gamma(n-2) & \cdots & 1
\end{pmatrix}; \hspace{1cm} (3.2)$$
\[
c(n) = \begin{pmatrix}
\gamma(1) \\
\gamma(2) \\
\gamma(3) \\
\vdots \\
\gamma(n+1)
\end{pmatrix};
\quad F(n) = \begin{pmatrix}
0 & \cdots & 0 & 0 & 1 \\
0 & \cdots & 0 & 1 & 0 \\
0 & \cdots & 0 & 1 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
1 & \cdots & 0 & 0 & 0
\end{pmatrix}. \quad (3.3)
\]

It is easy to notice that multiplying a vector with \( F(n) \) reverse the given vector.

As mentioned earlier, we are looking for a relation that gives the present from the past. With what we defined, we can then write

\[
\Gamma(n + 1) = \begin{pmatrix}
1 & c(n)' \\
c(n) & \Gamma(n)
\end{pmatrix}
= \begin{pmatrix}
\Gamma(n) & F(n)c(n) \\
c(n)'F(n) & 1
\end{pmatrix}, \quad (3.4)
\]

where the prime denotes vector transpose.

From this covariance matrix, we want to find out the conditional distribution of \( X_{n+1} \) given its past. By inverting the covariance matrix, it can be shown that for \( x \in \mathbb{R}^{n+1} \) and \( y \in \mathbb{R} \):

\[
(y \ x') \Gamma(n + 1)^{-1} \begin{pmatrix}
y \\
x
\end{pmatrix} = \frac{(y - d(n)'x)^2}{\sigma_n^2} + x'\Gamma(n)^{-1}x, \quad (3.5)
\]

where \( d(n) = \Gamma(n)^{-1}c(n) \). Therefore, we can deduce that the distribution of \( X_{n+1} \) given its
past have mean and variance as follow

\[
\mu_n = c(n) \Gamma(n)^{-1} \begin{pmatrix} X_n \\ \vdots \\ X_1 \\ X_0 \end{pmatrix}, \quad \sigma_n^2 = 1 - c(n) \Gamma(n)^{-1} c(n). 
\] (3.6)

The trick is to compute this recurrence efficiently. For this, we need to prevent matrix calculations at each step. This is why Hosking [18] proposes to compute \( d(n) \) recursively. [12] suggests a slightly different and faster version. An interested reader should consult these papers for more information. Finally, we note that these methods have complexity of order \( N^2 \).

One of the main characteristic of the Hosking method is that the technique is exact. This means the process it produces has exactly the desired Hurst parameter. Also, we can note that we do not have to specify the desired length of the sample before the simulation. It can be done on-the-fly. It is particularly useful if we want the simulation to stop at a random time.

### 3.1.2 Davies and Harte Method

In this method [11], we try to find a “square root” of the covariance matrix. By that we mean that we are looking for a square matrix \( G \), such that \( \Gamma = GG' \). Suppose we want a sample of size \( N \) and suppose that \( N \) is a power of 2, \( N = 2^g \), for some \( g \in \mathbb{N} \). In order to efficiently find a solution to this problem, the best way is to create a circulant matrix of size \( 2N = 2^{g+1} \) from the autocovariance matrix. If we call it \( C \), the circulant matrix is built as
follow

$$
\begin{pmatrix}
\gamma(0) & \gamma(1) & \cdots & \gamma(N-1) & 0 & \gamma(N-1) & \cdots & \gamma(2) & \gamma(1) \\
\gamma(1) & \gamma(2) & \cdots & \gamma(N-2) & \gamma(N-1) & 0 & \cdots & \gamma(3) & \gamma(2) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(N-1) & \gamma(N-2) & \cdots & \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(N-1) & 0 \\
0 & \gamma(N-1) & \cdots & \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(N-2) & \gamma(N-1) \\
\gamma(N-1) & 0 & \cdots & \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(1) & \gamma(0) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(1) & \gamma(2) & \cdots & 0 & \gamma(N-1) & \gamma(N-2) & \cdots & \gamma(1) & \gamma(0)
\end{pmatrix}
$$

A quick observation shows that the \(i\)th row is simply built by right-shifting the first row by \(i - 1\). The elements that are coming out at the right side of the vector just need to be reintroduced at the left side. Moreover, the matrix is symmetric. We observe that the upper left corner is the covariance matrix. It can be shown that for fractional Gaussian noise, the matrix is considered positive definite. Note also that if the desired sample size is not a power of 2, we need to fill the first row with more zeros.

Now that we have a circulant matrix formulation, we use one important theorem to achieve our goal. A circulant matrix has the property that it can be decomposed in the following fashion

$$
C = QAQ^*,
$$

(3.7)

where \(A\) is the diagonal matrix of the eigenvalues of \(C\), and \(Q\) is an unitary matrix defined
by

\[(Q)_{jk} = \frac{1}{\sqrt{2N}} \exp\left(-2\pi i \frac{jk}{2N}\right), \quad \text{for } j, k = 0, ..., 2N - 1, \quad (3.8)\]

\(i\) being the imaginary number. For a matrix \(Q\) to be unitary, we need to have \(QQ^* = I\), where \(Q^*\) is the complex conjugate transpose of \(Q\), which is easy to deduce. The eigenvalues are defined by

\[\lambda_k = \sum_{j=0}^{2N-1} r_j \exp\left(2\pi i \frac{jk}{2N}\right), \quad \text{for } j, k = 0, ..., 2N - 1, \quad (3.9)\]

where \(r_j\) is the \((j + 1)^{th}\) element of the circulant matrix first row.

Recall that the circulant matrix is symmetric and positive definite. Thus its eigenvalues are positive and real. We can state that a matrix with eigenvalues \(\sqrt{\lambda_1}, ..., \sqrt{\lambda_{2N-1}}\) and the same eigenvectors than \(C\) is also positive definite and real. Define the matrix \(S = QA^{1/2}Q^*\). Since \(Q\) is unitary, we can deduce that \(SS^* = SS' = C\), which is the representation we are looking for. Here \(S\) is the square root of the covariance matrix.

Starting from an i.i.d. vector \(V\), we need to simulate \(QA^{1/2}Q^*V\). This is done using the following three steps:

1. From (3.9), we first need to compute the eigenvalues. This can be solved efficiently using Fast Fourier Transform.

2. Compute \(W = Q^*V\).

3. Finally, solve \(Z = QA^{1/2}W\) using

\[Z_k = \frac{1}{\sqrt{2N}} \sum_{j=0}^{2N-1} \sqrt{\lambda_j} W_j \exp\left(-2\pi i \frac{jk}{2N}\right). \quad (3.10)\]
This can also be computed fast using FFT. By taking the first $N$ elements of $Z$, we obtain the fractional Gaussian noise.

Since the second part of $Z$ is built in a similar fashion, from the circulant matrix, we obtain a second fractional Gaussian noise. Nevertheless, the correlation between those two processes is different, such that we cannot link those two to obtain one bigger signal. Because of the FFT, the greatest advantage of Davies and Harte over Hosking is the speed. In effect, the complexity is then of order $N \log(N)$. We also note that this method is exact.

3.1.3 Wavelets-Based Synthesis

We describe here an approximate method to simulate the fractional Brownian motion. This is different from the preceding two methods.

If $B_H(t)$ is a fractional Brownian motion, then

$$d_{B_H(t)}(j, k) = \langle B_H(t), \psi_{j,k}(t) \rangle = \int \mathbb{R} B_H(t) \psi_{j,k}(t) dt$$

(3.11)

give the details obtained from the extension by the wavelet $\psi$ (see equations (2.13) and (2.16)). In [17], Flandrin studies the behavior of the variance of details and finds it to be $\sigma^2 2^{j(2H+1)}$ for some constant $\sigma^2 > 0$. If the chosen wavelet function has at least one vanishing moment, it can be checked that the coefficients are almost independent. In other words, the autocovariance of $d_{B_H(t)}$ are closed to zero. The Haar wavelets, which have zeroth-order vanishing moment, i.e. $\int t^k \psi(t) dt = 0$, for $k = 0$ hardly verify this property.

We then know the behavior of the details of a fractional Brownian motion. It is therefore somewhat possible to generate an alike process using (2.15). This has been accomplished by
Wornell [34]

\[
\tilde{B}_H(t) = \lim_{J \to \infty} \sum_{j=-\infty}^{J} \sum_{k \in \mathbb{Z}} \tilde{d}_{B_H}(t)(j,k)2^{-j/2}\psi(2^{-j}t - k),
\]  

(3.12)

where approximations of the details \( \tilde{d}_{B_H}(t) \) are independent Gaussian random variables with variance \( \sigma^22^{j(2H+1)} \). This ensures the properties the details should have. Wornell [34] shows that the generated process can be used as an approximation. It is further shown that the quality improves with the number of vanishing moments \( N \), but this is done at the cost of speed. Nevertheless, reference [2] argues that this method suffers from several drawbacks. The main one is the ignorance of the approximation at the coarsest level namely \( \text{approx}_{J} \). Wornell ignores it because it holds the low-frequency behavior of the sample. Ignoring this part is close to forgetting the long-range dependence of the process, which is obviously not desirable.

Another method has been suggested by in [2], originally proposed by Sellan. For this method, we use a white noise process. The fact is that the wavelet decomposition of this noise yields a set of uncorrelated coefficients

\[
W(t) = \sum_{k} \mu_k \phi(t - k) + \sum_{j \leq 0} \sum_{k} \lambda_{j,k} \psi_{j,k}(t),
\]  

(3.13)

where \( \mu_k \) and \( \lambda_{j,k} \) are samples of i.i.d. white Gaussian processes. It can be shown that we are able to generate ordinary Brownian motion process by integrating white noise. To obtain a fractional Brownian motion, we \textit{fractionally integrate} equation (3.13). \( D^{(-s)} \) is the fractional integration operator

\[
D^{(-s)} f(x) = \frac{1}{\Gamma(s)} \int_0^x (x - y)^{(s-1)} f(y)dy.
\]  

(3.14)
Sellan claims that this fractional integration with proper s parameter is equivalent to equation \([2.4]\). Therefore, we obtain

\[
B_H(t) = \sum_k \mu_k (D^{-s}) \phi_k(t - k) + \sum_{j \leq 0} \sum_k \lambda_{j,k} (D^{-s}) \psi_{j,k}(t) \tag{3.15}
\]

with \(s = H + 1/2\). Nevertheless, this form is not sufficient either. In fact, fractional integration of the wavelets destroys their orthonormality. It means that results can still be obtained but not efficiently. Sellan therefore uses one more modification to obtain

\[
B_H(t) = \sum_k b_H(k) \phi_{0,k}^{(s)}(t) + \sum_{j \leq 0} \sum_k \lambda_{j,k} 4^{-s} 2^{j} \psi_{j,k}^{(s)}(t), \tag{3.16}
\]

where \(\lambda_{j,k}\) are still i.i.d. Gaussian random variables, \(B_H(k)\) is a fractional process, known as Autoregressive Integrated Moving Average, and \(\phi^{(s)}\) and \(\psi^{(s)}\) are suitably defined fractional scaling and wavelet functions, derived from the original functions. The interested reader is invited to refer to [2] for further considerations. To conclude, we note that the samples generated following this original scheme exhibit too many high-frequency components. To circumvent this undesirable behavior Bardet et al. propose to downsample the obtained sample by a factor 10 [4].

Now, we have at our disposition several ways to produce time-series that exhibits long-range dependence with a known Hurst parameter. We can evaluate the performances of the estimators discussed in Section 3.2 with these signals. It is vital to note that the estimators are hardly exact. In fact, the main indicator of long-range dependence, i.e. the behavior of autocovariance, cannot be computed from a finite sample. Finite samples with long-range dependence do exhibit special characteristics. By analyzing such characteristics, we can evaluate \(H\).
3.2 Estimation Methods

This section considers three H-estimators. Our estimators use (a) aggregated variance method, (b) R/S method, and (c) wavelets method to evaluate $H$. We implemented the two first methods in Matlab. The Matlab routine for the wavelets method has been developed by Abry and Veitch [1].

In order to test the estimators, we generate, using the Davies and Harte method, a self-similar process of Hurst parameter $0.8$ and length $N = 2^{14} = 16384$.

3.2.1 Aggregated Variance Method

This method is based on the main property of the self-similarity, i.e. the equality in distribution of two scaled samples at different time scales (see equation (1.9)). First, we define the aggregated process $X^{(m)}$

$$X^{(m)}_k = \frac{1}{m} (X_{km} + \ldots + X_{(k+1)m-1})$$

for $k = 0, 1, \ldots$.

Using equation (1.9), we compare the variance of the original and aggregated value, i.e. $Var(X^{(m)}_k) = m^{2H-2}Var(X_k)$. We can then apply the following procedure to estimate $H$ [2]:

1. Aggregate the process in $M$ subseries of length $m$, with $m$ not too large. Typically, we should have, at least, $2 < m < N/2$. We obviously have $M = \lfloor N/m \rfloor$. The previous suggestion about the value of $m$ then ensures that we have a sufficient number of subseries. Next, we compute the mean of each aggregated process $\overline{X^{(m)}_1}, \overline{X^{(m)}_2}, \ldots, \overline{X^{(m)}_M}$ as well as the overall mean

$$\overline{X^{(m)}} = \frac{1}{M} \sum_{k=1}^{M} X^{(m)}_k. \tag{3.17}$$
2. For each chosen length of subseries $m$, we compute the sample variance of the sample means $X_k^{(m)}(k = 1, ..., M)$:

$$Var(X_k^{(m)}) = \frac{1}{M - 1} \sum_{k=1}^{m} (X_k^{(m)} - \overline{X}^{(m)})^2,$$

(3.18)

where $\hat{x}$ denote estimate of $x$.

3. Plot the log-log diagram of $Var(X_k^{(m)})$ versus $m$. This should yield a line with a slope of $2H - 2$. In practice, we fit a line through the generated points. We then have $H = slope/2 + 1$ as the Hurst parameter estimate.

It is important to note that equation (3.17) is biased in presence of non-zero correlation and therefore in presence of long-range dependence. For large $M$ the effect is somewhat decreased. Nevertheless, we have [12]:

$$E[Var(X_k^{(m)})] \sim Var(X_k^{(m)})[1 - CM^{2H-2}],$$

(3.19)

for some constant $C$ and $M \to \infty$. In case of long-range dependence, $C$ is positive. For small $M$, $Var(X_k^{(m)})$ is underestimated and leads to underestimation of $H$ as well. A small $M$ means a few samples to compute the variance. It might be wise not to take those values into accounts. Same can be said about large $M$. In effect, it means that we have a lot of small subseries. Some short-range dependence could appear between them, giving misleading information. Nevertheless, in the case where no short-range dependence is present, we achieve a good estimate of the variance. Note that processes such as Poisson have a slope of $-1$. In effect, their increments are independent and they have a Hurst parameter of $1/2$. 

34
Figure 3.1: The Aggregated Variance Method

Figure 3.1 shows the result of the log-log diagram of the variance plot. As illustrated, the approximated slope of the plot is $-0.46$. Therefore, the estimation of $H$ is $-0.46/2+1 = 0.77$. Although a little bit underestimated, which was expected, the value is pretty close to the actual Hurst parameter of 0.8.

3.2.2 Rescaled Adjusted Range Method (R/S)

The rescaled adjusted range or R/S method is one of the most popular methods to estimate long-range dependence. It was first introduced by Hurst [19], in the context of the flooding
of the Nile river, but then has been studied in numerous papers. Notice that to analyze signal \( X \), we virtually divide the original time series in \( K \) subsets of size \( k \). Obviously, we have \( k = \lfloor N/K \rfloor \). We define \( W_j \) to be the accumulation of \( X \) of length \( j \), i.e. \( W_j = \sum_{i=1}^{j} X_i \). \( W_j \) is also called \( j \)-aggregated series. For an arbitrary value \( r \), we define
\[
\mu_i(l, r) = W_{l+i} - W_l - \frac{i}{r} (W_{l+r} - W_l); \quad 0 \leq i \leq r.
\]
Use the rescaled adjusted range as
\[
R(l, r) = \max_{0 \leq i \leq r} \mu_i(l, r) - \min_{0 \leq i \leq r} \mu_i(l, r).
\] (3.20)
Considering the sample mean \( \overline{X}_{l,r} \) of a block of size \( r \) of \( X_{l+1}, \ldots, X_{l+r} \)
\[
\overline{X}_{l,r} = \frac{1}{r} \sum_{i=l+1}^{l+r} X_i,
\] (3.21)
we can define the sample deviation
\[
S(l, r) = \left\{ \frac{1}{r} \sum_{i=l+1}^{l+r} (X_i - \overline{X}_{l,r})^2 \right\}^{1/2}.
\] (3.22)
Naturally, \( S^2(l, k) \) is the sample variance. Now, restrict the starting point of our samples, i.e. \( l \), to be the starting point of the \( K \) virtual blocks of size \( k \). In other words, we have \( K \) possible starting points \( l_1, \ldots, l_K \). From there, the R/S statistics is computed as
\[
Q(l_i, r) = \frac{R(l_i, r)}{S(l_i, r)},
\] (3.23)
where \( i = 1, \ldots, K \) and \( r \) is the range. It is also called the lag. It is important to say that the only restriction on \( r \) is that \( l_i + r \leq N \). Also, note that it is easy to verify
\[
\mu_i(l, r) = \sum_{j=1}^{i} [X_{l+j} - \overline{X}_{l,r}] .
\] (3.24)
In practice, we perform the following steps:
1. Divide the time series in $K$ intervals of length $k = \lfloor N/K \rfloor$.

2. Compute the R/S statistics for the starting point of each block, for any possible value of $r$. Once again, it is important to notice that the only restriction on $r$ is that $l_i + r \leq N$. It means that for any range smaller than the length of a block, i.e. $r \leq k$, we have $K$ values of the R/S statistic. On the other hand, if $r > k$ we have less and less values available. In the extreme case, we only have one sample. If $r \geq N - N/R$, only the starting point $l_1 = 0$ satisfies the condition $l_i + r \leq N$. We notice that for large value of $r$, the samples become correlated since part of them are used in more than one computation of the R/S statistic.

3. We plot the logarithm of $Q(l_i, r)$ versus the logarithm of $r$. It is sometimes called the pox diagram. We may also take the average value of the R/S statistics, $Q(r) = \overline{Q(l_i, r)}$, $i = 1, ..., K$ before computing the logarithm. Statistically, the plot of log $Q(r)$ versus log $r$ should ultimately, for asymptotic behavior of $r$, be scattered randomly around a straight line with slope $H$. Thus, the slope of a least square regression line fitted to the points of the R/S plot gives an estimate of the Hurst parameter.

It is a little difficult to figure out why this method works. But we can come out with an intuitive interpretation of the R/S statistics, which could explain the result [12]. Recall we are considering fractional Gaussian noise, and consider the alternate value of $\mu_i(l, k)$ developed in equation (3.24). The mean value of a gaussian noise is approximatively zero. Using this as a fact, equation (3.24) is a cumulative sum of fractional Gaussian noise and $\mu_i(l, k)$ is the sample of size $r$ of a fractional Brownian motion process. Next, the rescaled adjusted range
computes the difference between the maximum and the minimum of the sample. Therefore, $R(l, r)$ is a measure of the dispersion and provide an estimation of the variance, which is proportional to $r^{2H}$. We divide by the deviation of the sample such that $R/S$ is proportional to $r^H$, giving the Hurst parameter.

Once again, practically, the smallest and largest values of $r$ should be ignored in the fitting process because of short-range dependence and lack of samples, respectively.

Figure 3.2 shows the results of the R/S analysis on our sample fraction Gaussian noise. Notice that we average the different values of $R/S$ for a same value of $r$. This gives a better
view on the alignment of the values. The value we found is 0.77, which is same as variance plot. It is underestimated but close to exact value of 0.8. With this value, there is no doubt that long-range dependence has been detected.

3.2.3 Abry-Veitch Estimator: Wavelets Method

Section 2.3 introduced the key concepts of wavelets expansion. We showed how wavelets could be used to generate an approximation of fractional Gaussian noise in Section 3.1.3. Here, we present an estimator of the Hurst parameter and we call this estimator as the Abry-Veitch (AV) or wavelets estimator [1]. Recall that wavelets decomposition is similar, in principle, to a Fourier transform. While Fourier analysis emphasizes the frequency behavior of the signal, wavelet coefficients possess scaling properties, because of the dilation of the bases. This explains why they are very convenient to analyze and characterize long-range dependence and self-similarity. We may say that the wavelet decomposition of such a signal transforms its autocorrelation. In effect, the LRD becomes a short-range dependence in the wavelets coefficient plane \{j, k\}, where j is the octave or level and k is the coefficient number. By analyzing the dependence of the details, we can thus derive the behavior of the original signal.

We note that the details at a given octave j, say \(d_X(j,.)\), form a stationary process with weak short range dependence. If \(n_j\) is the number of details at octave j, we have

\[
\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} |d_X(j,k)|^2.
\]

(3.25)

Where \(\mu_j\) is an unbiased estimator of the variance of the series containing details at level j. It can be seen as a medium to capture the second order behavior of the original process at level
Also, there exists no correlation between the different octaves. Since we are interested in the second order behavior of the signal, the evaluation of \( \mu_j \) could be a better option. From this, we produce a Logscale Diagram, which is a log-log plot of \( \mu_j \) as a function of \( j \). In fact, \[3\] shows that the expectation of (3.25) leads to a smooth expression, notably because of the long-range dependence. It is shown that

\[
\log \mu_j = (2 \hat{H} - 1)j + \hat{c}
\]

(3.26)

where \( \hat{c} \) is a constant which depends on the chosen wavelet. In other words, the estimator is actually giving the estimation of the parameter \( \hat{\alpha} \), since \( \hat{H} = (\hat{\alpha} + 1)/2 \).

This estimator has several useful and interesting characteristics. First, it can be checked that with a sufficient number of vanishing moments the estimator is unbiased even for short data sets. Second, the estimator allows to easily derive intervals of confidence. Third, Abry and Veitch \[1\] state that the choice of the wavelets function is irrelevant. Only the number of vanishing moments modifies the performances of the estimator. Thus, a very useful feature of the wavelet estimators is linked to the number of vanishing moments. In Section \[1.1.1\], we introduced the concept of trend to model non-stationarity. Most of the estimators are going to be affected by those perturbations. Nevertheless, the vanishing moments of a wavelet function are defined by equation \[2.17\]. In other words, it means that the mother wavelet is orthogonal to the space of polynomials of degree \( D \leq N - 1 \). If we recall the computation of the details of wavelet expansion \[2.13\], we see that the details of the analyzed signal do not contain any components from a polynomial trend as long as the number of vanishing moments is strictly greater than the degree of the polynomial trend. It means that for
Figure 3.3: The Abry-Veitch Estimator: Wavelets Method

\[ X_t = S_t + P_t, \]  
we have
\[ d_X(j, k) = d_S(j, k) \text{ for } N \geq D + 1, \tag{3.27} \]

where \( D \) is the degree of the polynomial trend \( P \) and \( N \) denotes the number of vanishing moments. Since the AV estimator is based on the behavior of the details, the estimation of the Hurst parameter should not be biased, provided that the number of vanishing moments has been adequately chosen. Further discussions and more details on computations are given in [1, 2, 3, 31].

Figure 3.3 displays the results produced by the AV estimator. As usual, we need to fit
a line through the points, while being careful about the first and last parts of the plots. As we can observe on the graph, octave 1-3 and octave 9-12 do not have the same slopes as that of the middle octave. This is why it is advisable to remove the first part of the graph. Because of the weighted regression performed with the fitting, removing the last octaves do not change the value of $H$. The estimator produces a value of 0.569 for $\alpha$ and, thus, a value of 0.785 for $H$. An interesting features of the AV method is that it provides an interval of confidence, which in this case is $[0.760, 0.809]$. Although the Hurst parameter is still underestimated, the estimate is more accurate, and the actual value of 0.8 sits in the interval of confidence.

3.3 Conclusion

In this chapter, we have presented three methods to simulate fractional Gaussian noise with a specific Hurst parameter. The first two methods are exact, while the third one using wavelets is approximate. These methods are utilized to derive a fractional Brownian motion process. We have also discussed three H-parameter estimators, namely, the aggregated variance, the rescaled adjusted range and the Abry-Veitch estimator. We have explained their principles and illustrated their characteristics. All these techniques are programmed using Matlab. Next chapter studies more carefully the performances of the estimators in both stationary and non-stationary data context.
Chapter 4

Simulations under Non-Stationarity

This chapter tests the estimators presented in Section 3.2 under different non-stationary conditions. Specifically, we evaluate the performance of the different methods under non-stationarity. Nevertheless, before being able to coherently discuss those results, it is important to verify the performance with the process alone. This gives us a good idea about the bias introduced by each method, which should also be present with non-stationary signals.

We consider simulated traffic for the studies in this chapter. Chapter 5 uses data from real traffic measurements. Therefore, in Section 4.1 we first analyze the bias of the different estimators. To do so, we generate several processes with a given Hurst parameter. Then, we compared the averaged results obtained by each estimator for different value of $H$. Next, Section 4.2 evaluates the bias of the approximate simulation by wavelets synthesis. The procedure is the same as in Section 4.1. Finally, we observe the performances of the estimator when the signal is mixed with a deterministic trend. Section 4.3 reviews (a) polynomial
trends, (b) power-law trends, (c) sine trends, and (d) level-shift.

4.1 Estimator Performance

As we just stated, this section is used to compare the performance of the estimators without any perturbations on the analyzed signal. This is why we test several realizations of a signal with a same Hurst parameter. We chose to test the values of \{0.5, 0.6, 0.7, 0.8, and 0.9\}. Recall that for $H = 1/2$, there is no long-range dependence, so we can simply use a Poisson process. The other signals are generated using the Davies and Harte method, which is both exact and pretty fast. Note that similar results have been obtained with the Hosking method. The results obtained for the different values of $H$ for our set of estimators are displayed in Figure 4.1.

The results are presented under the form of box plots for every value of $H$. Those box plots are centered on the average value found by the estimator. The upper and lower limits of the box represent the standard deviation of the obtained results. The blue box, on the left-hand side of each graph, represents the results of the wavelets method. The red box, at the center of each graph, and the black box, at the right-hand side of each graph, are linked to the aggregated variance method and the rescaled adjusted range, respectively. We now analyze the results for each consecutive value of the Hurst parameter.

1. In Figure 4.1(a), the generated signal is a Poisson process. Therefore, the Hurst parameter has a value of 0.5. At first peek, we observe that the best estimator is the wavelets based estimator. It has an average value of 0.5075 and a deviation of 0.0089. This means that most of the values produced by the estimator are very close to the
Figure 4.1: Performance of the Estimators for Different Values of the Hurst Parameter.
exact value of the process. On the other side, the aggregated variance method is not as accurate. The mean result is still good, at 0.4977, but the deviation is larger, meaning that a lot of the produced values are pretty far from the actual parameter. Finally, the worst case is met with the rescaled range analysis. While the deviation is pretty small, at 0.0099, the mean value is 0.544. That means that all the values obtained from the R/S interpretation are overestimated.

2. In Figure 4.1(b), we generate a fractional Gaussian noise with $H = 0.6$. Once again, the Abry-veitch estimator works well, with an accurate average and a minim deviation of 0.6042 and 0.0177, respectively. On the other hand, the performance of the variance method gets worse. Once again, the deviation is larger than all the other methods (0.0223), but, this time, the mean value is more remote from the exact value, at 0.5716. This value is underestimated, which we already note previously. In effect, we argued in Section 3.2.1 that in presence of long-range dependence, the estimate is underestimated. Finally, the rescaled range method keeps steady performance, as most of the values it produces are close to each other. Nevertheless, the computed values are still overestimated. The mean is 0.64 and the variance is 0.01603.

3. In Figure 4.1(c), the performance of the estimators are very interesting. The exact value that should be provided is 0.7. As usual, the wavelets methods is almost flawless, since its performance box sits tightly on the exact value. The mean is 0.7022 and the deviation is 0.0226. The variance method keeps very low performance. Not only its mean result, 0.6667, is not close to the exact value, but its deviation is still pretty large
at 0.0485. Conversely, for this Hurst parameter, the rescaled range analysis improves its performance dramatically. Not only the mean, 0.6993, is very close to the exact value, but also the deviation is very weak, at 0.0116.

4. In Figure 4.1(d), the value of $H$ that should be found is 0.8. One more time, the Abry-Veitch method has excellent results. The mean result is 0.797 and the standard deviation is just 0.0191. The variance method keeps its mediocre performance, as, again, the mean is largely underestimated, at 0.7493, and, moreover, the results are spread out, with a deviation of 0.0329. Finally, the R/S method has pretty bad performance. Its mean result is only 0.7673, which is an underestimation of the actual value. This is interesting to notice, since, previously, this method had a tendency to overestimate the Hurst parameter. Also, as before, the different results are very concentrated, as the deviation is only 0.0133.

5. In Figure 4.1(e), we simulate signals, which Hurst parameter is 0.9. As for all the other values of $H$, the wavelets method proves to be very accurate and convenient. The mean is 0.8954 and the deviation is 0.0249, both very good results. Also, as a constant through this test, the variance method provides poor performances. It importantly underestimates the actual value, with a mean result of 0.8123. Moreover, it is not seldom for the variance method to produce values a lot lower, since the deviation is 0.04304. Finally, the rescaled method still underestimates the exact answer, with a mean of 0.8253. With a deviation of 0.0177, most of the produced values are very close to each other.
From this test, it is very easy to conclude that the only constantly accurate estimator is the Abry-Veitch. Any given value is close to the exact parameter. Also, note that the box do not show the interval of confidence of the wavelets method, which is another asset of this method, from the accuracy point of view. Definitely, we do not want to trust the aggregated variance method, as its results can spread very far from the actual value. Finally, we can actually use the R/S analysis as a relatively accurate estimator. Although, most of the time, its value are biased, positively or negatively, the deviation is very weak for each Hurst parameter. This means that this bias is very steady, and thus, known, for a particular value of $H$. Therefore, it can be used to modify the result provided by the method, and thus, used to determine an acceptable estimate.

4.2 Wavelet Synthesis Performance

In this section, we evaluate the performance of the wavelet synthesis of fractional Brownian motion. From the results in Section 4.1, we know the bias of each estimator. We can then evaluate the bias due to the approximation of the process by analyzing the results obtained by the various estimators. Once again, we test the process for $H$ varying between 0.6 and 0.9, with a step of 0.1. We do not test for a Hurst parameter of 0.5, since we know that this represents short-range dependence. The values obtained by the estimators are presented in Figure 4.2.

The results displayed in Figure 4.2 are very similar to those when the analyzed signal was generated by an exact method. The aggregated variance still produce very spread out results, with most of them underestimating the simulated value significantly. The wavelets
method set of result is still very concentrated, but this time, most of the outputs overestimate the actual Hurst parameter. Finally, the rescaled range analysis has the same behavior than before. It always has a small standard deviation. Also, it overestimates $H$ for low values and underestimates it for high values. Nevertheless, it seems that the obtained values are closer in this case. The best way to compare the results of Figure 4.2 is to compare them to those of Figure 4.1. This is done in table 4.2.

In table 4.2, D stands for Davies and Harte method, while W stands for wavelets analysis.
Table 4.1: Estimate of Hurst Parameter for the Davies & Harte and the Wavelets method.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$H = 0.6$</td>
<td>0.6042</td>
<td>0.6162</td>
<td>0.5716</td>
<td>0.5812</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>$H = 0.7$</td>
<td>0.7022</td>
<td>0.7154</td>
<td>0.6667</td>
<td>0.6638</td>
<td>0.6993</td>
<td>0.7223</td>
</tr>
<tr>
<td>$H = 0.8$</td>
<td>0.7978</td>
<td>0.8143</td>
<td>0.7493</td>
<td>0.7457</td>
<td>0.7673</td>
<td>0.7846</td>
</tr>
<tr>
<td>$H = 0.9$</td>
<td>0.8954</td>
<td>0.9201</td>
<td>0.8123</td>
<td>0.8234</td>
<td>0.8253</td>
<td>0.8438</td>
</tr>
</tbody>
</table>

method. Also, wav, var, and R/S, are the shortened version of wavelets, variance, and rescaled range estimators, respectively. The values denote the average results obtained with different methods. For the wavelets estimator, the values are constantly overestimated. Also the bias is growing with the value of $H$. The same reasoning applies for the R/S analysis, but with a bias more constant. Finally, the aggregated variance mean is not much affected. It does not matter much, as the deviation is still large. From these results, we see that this approximation introduces a reasonable bias compared to the desired value of $H$. For the following simulations, we use the Harte and Davies method. Not only the results generated by this algorithm are exact, but also the algorithm complexity is comparable to the one of the wavelets synthesis. We point out that simulations with any other generator provides similar results when it comes to compare estimator performance under non-stationarity. In effect, we are interested in the difference between the estimates under stationary and non-stationary conditions, and not the difference with the actual value of $H$. 

50
4.3 Non-Stationary Signal: Trend Analysis

This section analyzes the performance of the estimators when the signal contains some sort of non-stationarity. As discussed in Section 1.1.1, we reduce the set of possible non-stationarity to a popular set of trends. This means that our original signal is mixed with another signal, which introduces the non-stationarity. For the original signal, we generate a fractional Gaussian noise, say $X$, of length $N = 2^{14} = 16384$, using the Harte and Davies method. The Hurst parameter is set to 0.8 as it is a value commonly encountered in the traffic network.

4.3.1 Trend 1: Linear Trend

The first case we simulate is a simple linear trend. In other word, a signal “$at + b$” is added to $X$. Both the single and the trended signals are illustrated in Figure 4.3. As it can be seen, the amplitude of the parasitic signal is quite small. This is why the two signals are pretty similar. Even a small perturbation can affect consequently the results provided by the estimators. Those results are presented in Figure 4.4.

Figure 4.4 organizes the data as follow. The left column shows the results for the original signal, while the right one displays the outputs for the trended signal. A row corresponds to an estimator. From top to bottom we present the aggregated variance, the rescaled range, and the wavelets method results. Since we choose to use the same scale for a same method, it is very easy to compare the performance of the estimators in the presence of a trend. The most dramatic difference is met with the aggregated variance method. When there is no perturbing signal, we estimate $H$ to be 0.775. But when we add the linear trend, the estimate goes to 0.89. If we refer to the Figure 4.1(d), we observe that there is no way that the variance
method should have a value in this neighborhood, even with its large deviation. Moreover, we know that in presence of long-range dependence, the estimate is mostly underestimated. The second row shows the loglog diagram for the R/S method. Contrarily to the variance method, we observe here a striking similarity between the two plots. The values found for the regression are exactly same and the Hurst parameter is estimated to be 0.78. Finally, the same conclusion can be drawn about the Abry-Veitch estimator. Both plots are very alike, and the difference between the two estimates is minimum (from 0.805 to 0.807).

4.3.2 Trend 2: Parabolic Trend

Now we consider a parabolic trend, which has the form “$at^2 + bt + c$”. This time, we use a trend with a large amplitude of several times the variance of the original signal. The two signals are illustrated on Figure 4.5.

Since the form and the magnitude of the trend have changed, we need to observe if the performance of the estimators are modified significantly. Figure 4.6 shows the results. Note
Figure 4.4: Performance of Estimators for a Linear Trend of Small Amplitude.
Figure 4.5: Analyzed Fractional Gaussian Noise: Parabolic Trend

that the plots are not always drawn at the same scale. The first estimation provides a value
of 0.73. For the second estimation, it is obvious that the points are not aligned anymore. We
can definitely not trust the result anymore. Moreover, we see that all the values comprise
between 2.81 and 2.84. This means that even if we try to fit a line through the points,
the slope is very close to 0. From this, it follows that the estimation of $H$ is 1, which is
obviously not correct. For the rescaled range method, we observe that the two plots are
this time slightly different. We can see that the points at the higher end are not strictly
aligned to those at the lower end. This leads to a slight bias on the estimation in the form
of an overestimation of the value. Actually, the value changes from 0.76 to 0.83. A method
that shows a dramatic difference is the wavelets method. The first thing one observes is the
difference between the two plots. While the first graph has an almost constant slope from
octave 3 to octave 9 (actually, because of interpolation and the weighted regression, one may
use octave 3 to 12 without much difference), the slopes of the consecutive octaves of the
Figure 4.6: Performance of Estimators for a Parabolic Trend of Large Amplitude.
The second graph are always growing. This is why the Abry-Veitch estimate value changes from 0.794 to 0.928. Note that if we try to limit our regression to the octave that have an almost constant slope (say octave 3 to 7) we obtain a value of 0.815. It is difficult to have a very high confidence on that value. But this does not mean that the Abry-Veitch method is not robust to non-stationarity trends. In effect, until now, we used the estimator with only one vanishing moment. In Section 3.2.3 we stressed that by choosing the adequate number of vanishing moments, we can eliminate any polynomial trends perturbing the original signal. We now proceed and increase the number of vanishing moments. Plots are shown in Figure 4.7.

The results obtained are 0.780 and 0.779. While those values are slightly biased, we obviously see that they do show that increasing the number of vanishing moments allow the Abry-Veitch estimator to become robust to the trend. Nevertheless, those results do not actually match the theoretical concept that was advanced by Abry and Veitch. They
argue that we need to have the number of vanishing moments strictly larger than the degree of the trend. Here, we have a trend of second degree, therefore, we should need at least three vanishing moments to estimate $H$ correctly. Nevertheless, two vanishing moments are sufficient. Actually, same could have been said from the results in Section 4.3.1. In that case, the trend was of degree one, and was solved without any problems with only one vanishing moment. More about the vanishing moments and their implications on the trend suppression is discussed in Section 4.3.3 below.

### 4.3.3 Trend 3: Polynomial Trend

Consider a trend, which is a general polynomial of the form $a_n t^n + a_{n-1} t^{n-1} + \ldots + a_1 t + a_0$. We choose $n$ to be 4. Also, this time, we choose a large magnitude for the trend. So huge, that it is actually pretty much unrealistic. The maximum magnitude of the trend is set to around $10^4$ times the variance of the original signal. Both the signal and the trend are depicted in Figure 4.8. A depiction of the trended signal would be useless here, since the variations of the signal would be almost invisible due to the magnitude of the trend.

We show the results provided by the estimators in Figure 4.9. As usual, the left column represent the original signal, and the right column shows the trended signal results. As we would have expected the aggregated variance does not perform any better than before. Its estimate changes from 0.785 to about 1, which is one more indication of the non-robustness of this method towards non-stationary trends. On the graph (c) and (d), we see, for the first time, the rescaled range analysis being highly affected by the trend. In effect, its estimate of $H$ starts at 0.77 and ends at 1. We discuss later whether this bias is due to the high-degree
polynomial, or to the huge magnitude of the trend. Finally, we discuss more closely the results provided by the Abry-Veitch estimator. The estimate of $H$ is 0.8 exactly for the original signal. The first thing we must notice is that the slope of all the octaves are aligned. This should be a good sign, as it means that the value of the Hurst parameter has been found similar for all the octaves. Nonetheless, the value found is 1.994 which is impossible. Recall that at this point we use only one vanishing moment. Also recall that the chosen trend is of the fourth degree.

Figure 4.10 shows the loglog plot for increasing number of vanishing moments. On graph (a), corresponding to 2 vanishing moments, we observe immediately that the octaves are not aligned anymore, as the first octaves have a much smaller slope. The value found is still irrelevant at 1.743. The smoothening of the slope carries on on graph (b) using a 3 vanishing moments wavelet. The smaller slope part now extends until the seventh octave, providing a value of $H$ of 1.030 more reasonable, but still highly biased. With 4 vanishing moments, as
Figure 4.9: Performance of Estimators for a Polynomial Trend of Large Amplitude.
presented on graph (c), the heavy sloped part almost disappear from the plot. Therefore, the $H$ estimate becomes a lot more accurate as it is evaluated to be 0.834. Finally, the graph (d) in Figure 4.10 shows the result if we use the Abry-Veitch estimator with 5 vanishing moments. This time, none of the octaves show a highly different behavior than the others. Almost all the octaves have the same slope, and the estimate now becomes very accurate at 0.801. The fact that all the slopes are now almost constant among the octaves is an excellent sign that we now have overcome every aspect of the non-stationarity. Moreover, if we observe the graph for 6 and 7 vanishing moments, we do not pinpoint major differences. Same can be said from their estimates which are 0.795 and 0.782. Therefore, for a very large trend magnitude, we are actually able to strictly verify the assumption of Abry and Veitch. We find that, if and only if $N \geq D + 1$, where $N$ is the number of vanishing moments and $D$ the degree of the trend, the trend is totally removed from the signal. Another observation that has to be made is that by increasing the number of vanishing moments, we are loosing some octaves. This means less data to analyze, and therefore, a smaller confidence on the obtained result.

Until now, we analyzed the estimators performance with three polynomial trends. In each experiment, we increased both the degree of the polynomial and the amplitude of the trend. We observe that the aggregated variance method could not handle any kind of non-stationarity. Further, we saw that with an adequate number of vanishing moments the wavelets method can always ignore the trend part of the signal. In this last case, for the R/S analysis, the magnitude of the signal might be an important factor. This is why we now
(a) 2 vanishing moment
(b) 3 vanishing moments
(c) 4 vanishing moments
(d) 5 vanishing moments

(c) 6 vanishing moments
(d) 7 vanishing moments

Figure 4.10: Wavelets Estimator for Varying Vanishing Moments: Large Amplitude.
analyze the same polynomial signal, but with a smaller amplitude

We consider the same fractional Gaussian noise. Results provided by the estimators on the original signal are depicted in Figure 4.9. Now, we add to the signal a trend of degree four but with only an amplitude of about 30 times the variance. The resulting signal is shown in Figure 4.11(a). With this new perturbed signal, we observe how the estimators perform. Results are shown in Figure 4.11(b), (c), and (d).

Mostly, results do not diverge from their usual forms. The variance plot is totally biased...
by the non-stationarity and estimate $H$ around 1. Also, while not as affected, the R/S analysis value is highly overestimated compared to the one originally issued (0.88 from 0.77).

What it is interesting to note is that in this case, the rescaled analysis bias is bigger than for the case studied in Section 4.3.2. In effect, if we consider Figure 4.6, we see that the estimate change from 0.78 to 0.83. In that case, the trend was a second degree polynomial with an amplitude of about 50 times the variance. Therefore, it means that the bias met by the rescaled range is affected both by the complexity of the polynomial, and its magnitude.

Finally, we analyze the results of the Abry-Veitch estimator with one vanishing moment. This time, we observe that the loglog plot does not produce a straight line as it does in Figure 4.9. The plot is actually closer to the one in graph 4.10(b) for 3 vanishing moments. The Hurst parameter is then evaluated to be 0.998. As usual, we now increase the number of vanishing moments and determine the improvement, if any, on the estimate.

Figure 4.12 shows the plots for increasing number of vanishing moments. This time, the heavy slope of the high octaves are quickly removed as with only 2 vanishing moments all the octaves are almost aligned. Actually, with only two vanishing moments, we already produce an estimate of 0.830. With three vanishing moments, the value becomes even more accurate at 0.792. The fourth moment brings few difference in the plot, and even less in the estimate, where we get 0.792. This is a very important observation. In effect, we are here in a deterministic environment. Nevertheless, if we were in a random environment, we would conclude, following the theory from the Abry-Veitch estimator, that the non-stationarity is of degree three, since adding a vanishing moment does not improve the estimate anymore.
Figure 4.12: Wavelets Estimator for Varying Vanishing Moments: Moderate Polynomial.
This would lead to a bad hypothesis for someone who would try to model the non-stationarity. If we add a fifth moment, we find the most accurate value, at 0.801. Nonetheless, the only way to guess that the optimal number of moments is five is by observing the loglog plot. In effect, we easily point out that from five moments to seven, the appearance of the plot has almost no modification. This is a clue that the degree of the trend is four. Nevertheless, it is not always easy to make such observations. Note that for 6 and 7 moments, the estimate of $H$ are 0.795 and 0.782, respectively.

We also tried with a polynomial having a smaller amplitude. In such a situation, the values provided by the variance method and the R/S analysis are, respectively, 1 and 0.77. The R/S analysis is thus unaffected. For the wavelets method, we get, successively: 0.827, 0.808, 0.792, 0.792, 0.801, 0.795, and 0.782. If we were in an unknown environment, it would have been very difficult to determine if there is even a trend is that case. In effect, the first value outputted by the estimator is already quite accurate. Actually, it is more accurate any values provided by other methods. Once again, the graph do show some similarities after the five moments estimator, but here, even more than before, the variations are very slight.

We can draw some useful conclusions from these discussions on polynomial trends. First, the aggregated variance is never to be trusted in an environment where there could be some non-stationarity. Actually, the values provided are so absurd that no one could believe them anyway. One useful utilization of the aggregated variance method would be to detect the non-stationarity environment as obvious from its absurd values. Second, the rescaled range analysis undergoes a bias depending on the complexity and the magnitude of the trend.
Hence, the R/S method should be used only when we know that the eventual non-stationarity is not too large in size and/or quite simple. Finally, results showed that, once again, the wavelets method is accurate, if we choose sufficient vanishing moments. Nevertheless, if used to define the order of the non-stationarity, we saw that the Abry-Veitch estimator was less accurate. The problem is that even with smaller vanishing moments than needed, this estimator is robust enough to derive an accurate estimate of $H$. This is why it is very tough to model the trend with such a method.

### 4.3.4 Trend 4: Increasing Power-Law

We now consider trends that cannot be modeled by a polynomial. The first such trend that we are going to treat is the increasing power-law, which has the form “$at^\alpha$”. For this simulation, we choose $\alpha$ to be $3/2$. We adopt the same layout of presentation than in the other sections with the original signal and its trended version displayed in Figure 4.13, while the results provided by the estimators are displayed in Figure 4.14.
Figure 4.14: Performance of Estimators for an Increasing Power-Law Trend.
While we change the trend types in each section, the performance of the variance method do not improve. As usual the value found after perturbation is still around one. Despite the fact that the amplitude of the trend is pretty large, the R/S analysis gives a very good estimate with a minimal bias. For the wavelets, as usual, we can see that the octaves at the end of the diagram are a lot steeper. Therefore, once again, the number of vanishing moments help us go around this issue. The loglog plots are presented in Figure 4.15.

Once again, the robustness of the Abry-Veitch estimator is proved by the results. The original estimate is 0.785. Then an increasing power-law is added to the signal. By increasing the number of vanishing moments, we obtain 0.860, 0.771, 0.781, 0.774, 0.775, 0.767, and 0.771. The values obtained are very close to each other. Same can be said about the loglog plots. The only plot that looks slightly different is the graph (a) which still has a heavy slope on the last octave. Nonetheless, it is very difficult to determine from here what is the optimal number of vanishing moments, if we do not know the original value. Our conclusion is then similar as before. We acknowledge the fact that a sufficient number of vanishing moments can efficiently analyze a trended signal. This property is true even when the trend is not polynomial. Nevertheless, an optimal number of vanishing moments is difficult to determine.

4.3.5 Trend 5: Decreasing Power-Law

In this section we are interested in the decreasing power-law trend. From the previous sections, we can definitely conclude that the aggregated variance method is not suited for any, even slightly, perturbed signal. Therefore, we do not discuss further its results in the following sections. On the other side, the rescaled range and the wavelets estimator
Figure 4.15: Wavelets Estimator for Varying Vanishing Moments: Increasing Power-Law.
Figure 4.16: Analyzed Fractional Gaussian Noise and Decreasing Power-Law Trend

(a) Moderate decreasing power law trend  (b) Huge decreasing power-law trend

have shown some good signs of robustness against non-stationarity. We observe them in
more details. Since we know that the amplitude of the trend can also have an important
influence, we analyze here two decreasing power law, respectively, with a moderate and a
huge magnitude. The power-law is of the form $at^{-1/4}$. The two resulting trended signals are
displayed in Figure 4.16.

In Figure 4.17, we expose the results of the two estimators analyzing a moderately trended
signal. With such an amplitude the rescaled range is barely affected, as its estimate stays
constant at 0.77. The results are pretty much the same for the Abry-Veitch estimator. The
original estimation is 0.798. Even with only one vanishing moment, the estimated value of
$H$ is very accurate with 0.806. Increasing the number of vanishing moments do not improves
much this approximation, as the estimates are, successively, 0.806, 0.792, and 0.792. Also,
the loglog plots all have a similar appearance. In this case, one might be the optimal number
of vanishing moments.
Figure 4.17: Performance of Estimators for a Moderate Decreasing Power-Law Trend.
In Figure 4.18, we show the results corresponding to the signal referenced in Figure 4.16(b). In this case, the trend is extremely large, such that the variation of the original signal cannot even be distinguished anymore. Under the influence of such a large amplitude, the R/S analysis does not provide an accurate estimate anymore. While the original value of $H$ has been estimated to 0.76, the trended signal estimate is 1. As we know, this value is absurd, and confirm our previous observations that a huge amplitude in trend makes the rescaled range method ineffective. Without the non stationarity, the Abry-Veitch estimator found the Hurst parameter to be 0.794. With the trend, it has several troubles as its loglog plot is a straight line which slope provides an estimate of 1.974. We therefore increase the number of vanishing moments. Quite surprisingly, with two and three moments, the correction on the slope of the different octaves is not effective at all. It actually yields estimates of 1.799 and 1.839, respectively. This situation is different from our previous tests, where increasing the number of vanishing moments gradually increased the quality of the estimate and the appearance of the plot. We thus continue to increase the number of vanishing moments to see if improvements finally shows up.

The results for more vanishing moment are presented in Figure 4.19. We start with four vanishing moments, and we try all the way to nine vanishing moments. This time, we see the usual improvement shown when we increase the number of vanishing moments. In effect, the first octaves gradually have a smaller slope. The number of octaves who has the same slope continuously increase, until they all have almost the same slope (graph (e) and (f)). The Hurst parameter estimate also evolves in the good direction, as we find,
Figure 4.18: Performance of Estimators for a Huge Decreasing Power-Law Trend.
successively: 0.980, 0.833, 0.804, 0.796, 0.799, and 0.801. Once again, it is very tough to
determine which number of vanishing moments is optimal. The closest value is given when
we use seven moments. Nevertheless, from six moments to nine, all the values are very
close and almost indiscernible. The interesting observation we can make here, is that the
Abry-Veitch estimator needs a large number of vanishing moments to be able to analyze the
signal correctly. In effect, before seeing any improvements, we observed that the estimates of
the Hurst parameter stayed pretty far from the actual value and that the loglog plots of the
method do not show any sign of improvement. With the first experiments, we showed that
the AV estimator is able to handle polynomial trends. This is because of the construction of
the details of the signal and the number of vanishing moments. Here, we observe that the
wavelets method is also robust to non-polynomial trends. This can be explained by the fact
that the latter can be approximated by a set of polynomials.

4.3.6 Trend 6: Sinusoidal Trend

This section analyzes the sine function trend, which has the form \( a \sin 2\pi ft \). This trend
introduces periodicity in the analyzed signal. Obviously, this is a tricky threat for the
estimators. Therefore, two parameters can now be used to prevent a correct estimation of
the Hurst parameter: the amplitude and the period. From our previous tests, we know
the effect of the amplitude on the estimators. We therefore concentrate our effort on the
influence of the number of periods in the signal. In Figure 4.20(a), the sine function repeats
4 times, while in Figure 4.20(b), the function repeats 16 times.

We first study the case of four periods in the signal. Figure 4.21 presents the plots drawn
Figure 4.19: Wavelets Estimator for Varying Vanishing Moments: Decreasing Law.
from each estimator. Figure 4.21(b) shows that, in that situation, the rescaled range value is affected, as the estimate goes from 0.77 to 0.84. On the other hand, the Abry-Veitch estimator is seriously affected, giving an estimate of 0.900 (the original estimate is 0.794), and showing a loglog plot with consequent difference in slope among the octaves. Note that the first estimate was 0.794. Surprisingly, even with an increased number of vanishing moments, the slopes of the end-octaves stay steep for a while. Nevertheless, the estimate is a lot more accurate with, successively, 0.816 and 0.795. If we continue to add some vanishing moments to the mother wavelet, the slope of all the octaves finally align, but the value is not as accurate as before, with, successively, 0.777, 0.771, 0.774, and 0.761. However, those values are pretty close and provide a good estimate of the Hurst parameter.

We now consider the case of a sine function with a smaller period. Thus, the sine appears more often in the analyzed signal, actually sixteen times. The situation is depicted in Figure 4.20(b). The results provided by the estimators are shown in Figure 4.22. The first graph
Figure 4.21: Performance of Estimators for a 4-Sine Trend.
shows that the R/S analysis is now highly biased by the presence of this periodicity in the
signal. However, we also observe that the loglog plot do not provide a straight line anymore,
but rather, it seems that there are two sets of points that have different slope. Considering
only the first set, we fit a line through the points. This is illustrated in the graph (b). This
time, the value we find for the estimate of \( H \) is better as we find 0.82. The results for
the Abry-Veitch estimator are even more interesting. Note that the estimate of the original
process is 0.801. In effect, from one part, the value provided by the method are somewhat
close to the actual Hurst parameter. In effect, we find, successively, 1.062, 0.949, 0.901, and
0.856. Nevertheless, it is not as accurate as it used to be in our previous tests. Also, we see
on the loglog plots that the wavelets estimator has trouble to find a common slope for every
octaves, which question the reliability of the approximations.

This is why we continue to increase the number of vanishing moments. The results are
shown in Figure 4.23. Some pretty interesting observations can be made from the plots
displayed. First, we note that a negative slope reappears at the last octave. This negative
slope was present in the graph (c), (d), and (e) of Figure 4.22. But with four vanishing
moments, all the slopes were positive. We would have think that this trend would continue
with more vanishing moments. But it turns out that we need at least eight vanishing moments
to meet this behavior again. Same could be said from the estimate of \( H \). Values, while
going towards the actual value, are not as accurate as they used to be. Actually, we have,
successively, 0.857, 0.849, 0.829, 0.806, 0.821, and 0.814. The closest value is provided with
eight moments, which is also the number of moments from which the negative slope disappear

78
Figure 4.22: Performance of Estimators for a 16-Sine Trend.
Figure 4.23: Wavelets Estimator for Varying Vanishing Moments: 16-Sine Trend.
in the logscale diagram.

4.3.7 Trend 7: Level Shift

For the last study, we consider the level shift. The level shift is a relatively quick and large change in the mean of the process. This trend is interesting because it is a very good approximation of what happen in the network, for example, at the peak hours. We see also that the results provided by the estimators are pretty interesting. We study three different level shifts in term of amplitude. The first test is done with a level shift of two times the variance of the original signal. The second and third level shift represent jumps of, respectively, five and ten times the variance. Those three trended signals are shown in Figure 4.24.

In this section, we organize the results differently that we used to. We first study the performance of the rescaled range analysis with the three different trends, and then we analyze the results produced by the Abry-Veitch estimator.

We first analyze the performance of the rescaled range analysis. Note that results on the
original signal are depicted in Figure 4.25(a), while those for the different level shift are in graph (b), (c), and (d). R/S method provides a value of 0.74 for the original signal. For the trended signals, the R/S analysis shows to be remarkably robust. This is true for all the tested amplitude. The bias stay always very little, as the most different estimate is at 0.75, which is very close to the original value found by this estimator.

Now, we analyze the case of the wavelets performance. We know, by theory and by our previous experiments, how big the number of vanishing moments in the mother wavelet is.
This is why we present the plots for varying number of vanishing moments for each amplitude of the level shift. In this case, we see that the observation of the graph turns out to be very interesting. First, note that the original Hurst parameter has been estimated at 0.781. The different plots outputted by the estimator are displayed in Figure 4.26, 4.27, and 4.28. Also, the estimates provided in each case are presented in table 4.2.

We first analyze the results of table 4.2. For a jump of two times the variance, we see that the Abry-Veitch estimator is pretty robust. When we increased the number of vanishing moments, the estimate becomes more accurate. This cannot be said about the next level shifts. For an amplitude of five time the variance, there is a consequent bias that prevents
Figure 4.26: Wavelets Estimator: 3-Level Shift.

the estimate to be accurate. Increasing the moments do improve the result. But even with a certain number of moments, the estimates stay pretty far from the actual value. Same can be said about the jump of magnitude ten. If we would have increased the jump even more, the estimate would be even more inaccurate, reaching inadequate values, even with a lot of vanishing moments. This means that the Abry-Veitch estimator is very sensitive to the level-shift. Therefore, in the potential presence of such a trend in the analysis of the
process, the measure provided by this estimator must be taken with precaution. This is very important to notice as we mentioned that the level shift is a good approximation of many scenarios met in real cases (One typical example being denial of service (DOS) attack). But the most important observation is done by examining Figures 4.26, 4.27, and 4.28. In effect, even a quick glance allows us to see that the evolution of the plots for increasing moments is very similar for each value of the jump. With few vanishing moments, there are some
irregularities between the slopes of the octaves. With many of them, the slopes tend to align, except for the last octave. But the evolution is same for every amplitude. However, when the amplitude is small, the estimate is accurate. This means that it is difficult to detect that the estimator has been biased by only observing the provided results. In the networking context, this flaw is a very important issues. Level shifts are very common in Internet traffic. Thus, one must be very cautious when we use the AV method on real traffic.
4.4 Conclusion

In this chapter, we used the tools and concepts we introduced in Chapter 3. First, we analyzed the performance of the estimators for different Hurst parameter. We found that the aggregated variance method performs very poorly. Contrary to this, the Abry-Veitch estimator is both accurate and constant. The R/S analysis is not as accurate, but has a constant bias. It can be used to approximately estimate $H$. This chapter then studied a known signal to which is added different kinds of trends. The main trend categories are polynomial, power-law, sine, and level shift. We applied the estimators to each of those situations and observed the outcomes. The variance method performs very poorly. It cannot handle any trends. Its only use would be to actually confirm the presence of a trend in the analyzed signal. Next, we observed that the Abry-Veitch estimator is particularly robust, even for very complex and huge trend, especially when we increase the number of vanishing moments. We notice two important flaws to this method. First, it cannot be efficiently used as a trend estimator. In effect, we see that such an evaluation can be accurately made only if the amplitude of the signal is huge. Second, we notice that this estimator is heavily affected by level shifts. Also, we see that this bias is difficultly noticeable on the plots. This is problematic as level shifts are a commonly known trend. Finally, we observe that the rescaled range estimator is affected by the complexity and the amplitude of the trend. It means that for simple and reasonable trends, the R/S analysis can be considered as accurate. In case of level shift, the R/S analysis performs actually better than the wavelets estimator.
Chapter 5

Application on Real Data

In this chapter, we apply the techniques previously studied on some real network activity data. Our goal is manyfold. First, we want to check if such network traffic exhibit long-range dependence, which would be detected by the different methods. Second, we want to observe any possible non-stationarity in the signal. Third, we plan to check if the results we found with the various techniques in Chapter 4 match the results we obtain on those data.

5.1 Data Background

The TCP data was collected from Penn State University’s operational network. Traffic was passively collected at a subnet firewall (see Figure 5.1) using tcpdump. Each sample was of a varying length of time with an average length of 8 hours. Recorded information included only IP header information (i.e. protocol, port...) and arrival timestamp. From each recording, a time-series was formed by counting the number of packets within consecutive, uniformly sized, time intervals. The time interval was one second. We were in possession of those
5.2 Real Data Analysis

We use the interarrivals packets time pictured in figure 5.2. The measurements have been taken on a period of one day or 86400 seconds. When the set is observed as a whole, we mainly see the day cycle, which appears as a big level shift. It starts around 7am and stop around 4pm. We can also see that after this cycle, there are two smaller, in size and time, jumps. Obviously, those cycles are non-stationarity that disturb an accurate estimation of the Hurst parameter. Also, one surprising observation we make concerns those numerous very high peak, we find highly above the mean of the signal. Those represent time when a really important number of packets has been received in a particular second. We see that it actually arrives quite often, even in the night cycle.

Now, we analyze the signal with the tools introduced previously. First, we apply the aggregated variance method. We already noticed the non-stationarity in the signal. Therefore, we do not expect the aggregated variance to give a correct estimate. The plot is shown in figure 5.3. As we can see, once again, this method is totally biased and provide an estima-
tion of $H$ around 0.95. This is a large value compared to what is normally found for network traffic. This is one more hint about the presence of non-stationarity in the analyzed signal.

Next, we apply the rescaled adjusted range method. In this case, we are not sure what to expect. First, we saw that the provided estimate, even in ideal conditions, is partly biased in most of the cases. Nevertheless, this bias is almost constant and approximately known for all the values $H$. Also, we saw that the R/S analysis is robust to non-stationarity if its amplitude is moderate. Since it is difficult to set a value where this method becomes biased, it is difficult to know in advance if the results will be good or not. The loglog plot is presented in figure 5.4. The estimate provided is then 0.77. A few hints here make us think that this value is quite accurate, at least for a value provided by the rescaled range. The
first thing is that in all the simulations we performed, the estimate was never decreased but always either unaffected or increased. Therefore, 0.77 is indeed, either an augmented value or the correct value. With the bias existing in the R/S analysis, we may think that the actual Hurst parameter ranges around 0.8, which is the usual value found for network traffic. This is another reason why we believe the rescaled range estimate is almost unaffected. Therefore, in this case, the R/S analysis is really robust to the present non-stationarities.

Finally, we present the results outputted by the Abry-Veitch estimator. From our simulations, this method is the most robust of all. We therefore expect the estimates to be
very accurate. The values provided are displayed in Table 5.1. \( N \) is the number of vanishing moments, and \( H \) is the estimate of the Hurst parameter.

**Table 5.1: Real Data: Abry-Veitch Estimator**

<table>
<thead>
<tr>
<th>( N )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H   )</td>
<td>0.995</td>
<td>1.008</td>
<td>0.997</td>
<td>0.980</td>
<td>0.984</td>
<td>1.003</td>
<td>0.970</td>
<td>1.001</td>
<td>0.989</td>
</tr>
</tbody>
</table>

Conversely on what we expected, the values here are all very high. Those values are
even worse than the estimate from the aggregated variance method. Moreover, increasing the number of vanishing moments does not improve the estimate. The only case where the wavelets estimator performed poorly is with a level shift. Therefore, we may think that the day cycle is viewed as a huge level shift by this estimator. The presence of this level shift might obviously disturb the Abry-Veitch estimator, leading to an incoherent value. But a closer look to the signal might gives us a better understanding of why the wavelets estimator fails in this case. We extract several part of the traffic, to take a more closer look on the traffic. Those parts of the signal are displayed in Figures 5.5, 5.6, and 5.7. An obvious observation is that we find a lot of level shifts in the signal, in all the extracted parts. Since we found that the more robust estimator in regard of this kind of non-stationarity is the R/S analysis, it is normal to find such an accurate value. On the other hand, the Abry-Veitch estimator is not robust at all to this kind of non-stationarity. Therefore, it provides values that are highly biased. Also, the daylight cycle looks much more like a second degree trend than an actual level shift. Since this trend does not have a huge amplitude, the rescaled range analysis handles it smoothly (see simulation in section 4.3.2). On the other hand, the Abry-Veitch estimator is probably not affected by the day cycle, but more by all the small sudden jumps in the traffic. Another reason why the rescaled range analysis perform so much better in that case than the wavelets estimator might be the length of the studied signal. In effect, by construction, the longer the signal is, the better the R/S estimate is. On the other hand, [1] claims that its estimator is unbiased, no matter the length of the studied signal. Therefore, only the accuracy of the R/S estimate is improved.
5.3 Conclusion

In this chapter, we analyzed some data from a real network. We could have repeated this study for different data of the same network, but it would have been pointless. In effect, all the data exhibits a similar pattern, and therefore, all the results provided by the estimators are comparable.

The main point that we show by analyzing this real data is that our conclusions in Chapter 4 are verified. In fact, we observe that in the presence of level shifts, the Abry-Veitch estimator has really poor performance. It can no longer be trusted in this context. Also, in that case, increasing the number of vanishing moments does not improve the estimate. On the other hand, we notice that the rescaled range analysis provides an accurate estimate.
This can be explained by two reasons. First, we showed that the R/S analysis is very robust to level shifts. It is therefore normal to find here that the value is pretty much unaffected. Second, the length of the data is important. Obviously, this improves the accuracy of the R/S analysis. Chapter 6 concludes our work and presents some potential future directions for the extension of the present work.
Figure 5.7: Traffic Trace: Zoom on End
Chapter 6

Conclusions and Future Work

The objective of this thesis was to characterize and model the network traffic behavior. First, we stated that the traffic was characterized by two important concepts: long-range dependence and self-similarity. Long-range dependence means that the past always has some influence on the present. Self-similarity assumes that two samples of the signal, viewed at different scales, look alike. By using these concepts, traffic modeling can be reduced to the estimation of Hurst parameter. Numerous methods exist to perform this estimation. Most of them are tested under the stationary process hypothesis. For short samples, stationarity might hold. But, this is hardly the case for larger samples. As an example, for network traffic, the day cycle is non-stationary. Therefore, the thesis aimed to test the robustness of the estimators under non-stationarity conditions.

In Chapter 1, we first introduced stationarity and non-stationarity concepts. We observed that the latter was equivalent to a popular set of trends. This chapter also described the
notions of long-range dependence and self-similarity in random process.

Chapter 2 presented two processes, namely, the fractional Gaussian noise and the fractional Brownian motion. They are used to model the network traffic. We introduced the concept of wavelets too. Finally, this chapter provided a brief overview on the origins of self-similarity in the network traffic.

Chapter 3 developed the tools used to estimate the Hurst parameter. We started by defining three methods to build fractional Gaussian noise for a given value of $H$. Next, we introduced three estimators: the aggregated variance, the rescaled range analysis, and the wavelets estimator. Using Matlab implementation the chapter has demonstrated the effectiveness of noise simulations and $H$-estimators.

Chapter 4 tested the tools developed in Chapter 3. First, we evaluated the bias of each estimator by testing them repeatedly on processes with a known Hurst parameter. We applied them for several particular values of $H$. Second, we added to our stochastic signal various kinds of trends in order to study the reaction of the estimators in the presence of non-stationarity. We analyzed mainly four categories of trends. Such as polynomial trends, power-law trends, sinusoidal trends, and level shifts. We saw that depending on a trend, the behavior of the estimator could change dramatically. We also observed that the amplitude of the trend have an influence on the estimators. From there, we derived a few conclusions. They are: (a) The aggregated variance could not handle any kind of non-stationarity. (b) Wavelets method performed well for almost every trend, even with large amplitude. Nevertheless, it was highly biased in presence of level shift and the estimation
of a trend was accurate only for a limited number of cases. (c) Finally, the rescaled range analysis performances were dependent on the complexity and the amplitude of the trend. Also, the R/S analysis was more robust to the level shift than the wavelets method.

In Chapter 5, we have applied the estimators on data from a real network. The trace was long and contained several non-stationarity. The primary traces belong to the day cycle. Results, obtained in Chapter 4, were re-confirmed. In effect, the variance method performed very poorly, the R/S analysis seemed to be accurate, and the Abry-Veitch estimator was highly biased. By taking a closer look on the time-series, we were able to understand the failure of the AV estimator. In effect, we found numerous level shifts in the data, which were perturbing the estimation by wavelets.

Note that it would be dangerous to rely only on one single method to evaluate the Hurst parameter. Under stationary conditions, each method performs relatively well. The same is not true if some kind of non-stationarity is present. Also, depending on the trend, each estimator has its own behavior. Particularly, we do not want to use the aggregated variance method. Under stationary conditions, this estimator showed a lowest performance. In presence of any kind of non-stationarity, its estimate was found to be highly biased. This method, however, has an important application as it helps confirm the presence of non-stationarity. We noticed that the rescaled range analysis is originally biased. The bias is almost constant for a given value of $H$. We can, thus, determine an approximate value of $H$ from the provided estimate. We also noticed that the complexity and the amplitude of the trend could modify the estimate as well. However, the R/S method was robust to level
shift. This property is really useful in the networking context as it is a common trend. The observation explains R/S method’s most realistic value for the real data estimation. Considered as the best method by many papers, we showed that the wavelets method had several flaws. First, the estimation of a trend by evaluating the optimal number of vanishing moments worked only for very large amplitudes. In other cases, the method is robust enough to handle the non-stationarity with a “sub-optimal” number of vanishing moments. Also, we observed that wavelets method was robust even when the amplitude was large. This statement is true only if the trend denotes a smooth trend. In case of the level shift, even a trend with moderate amplitude could perturb the estimator.

In the following, we provide some future directions for the work. We made it clear that relying on only one method to determine the Hurst parameter would be unreliable. Therefore, it might be interesting to try other estimators. A number of methods exist. They include absolute moments, discrete variations, Higuchi’s approach, Periodogram, and Whittle’s method. While these methods are studied considering stationary conditions, they need to be tested under different form of trends. By observing their behavior, one would have more flexibility to interpret results.

The orthogonal wavelets method has some drawbacks too. Because of dilation and translation, they are ideally suited to analyze self-similar processes. We showed that for a smooth trend, the non-stationarity can be virtually eliminated. Also, we found that the amplitude of the trend did not affect the wavelets estimator. It is, though, important to find out why level shifts introduce a bias in the estimation. Finally, Abry and Veitch notice that the
choice of the orthogonal wavelet is unimportant. Only the number of vanishing moments has an influence on the estimation. We might need to check if other particular wavelets would not lead to a better granularity to estimate the degree of a trend. Also, there may be some wavelets that would be able to handle a level shift. The thesis may utilize another kind of wavelets such as the biorthogonal wavelets. The biorthogonal wavelets are considered for JPEG-2000 standard. A biorthogonal wavelet requires two set of coefficients: one for the wavelets decomposition and the other for the signal reconstruction. The orthogonal wavelets use the same coefficients for both operations. One of the advantage of biorthogonal wavelets is that they are shorter. Thus, they are faster and can use more vanishing moments. They are, generally, constructed using lifting steps: split, prediction, and update. It is entirely spatial and exploit the correlation structure present in the signal. The correlation structure is typically local in space (time) and frequency. Neighboring samples and frequencies are more correlated than ones that are apart. Predictors can be based on subdivision methods, used extensively in Computer Aided Graphics Design to generate curves and surfaces. Spline methods fall into this category. A properly designed predictor based on the traffic data can offer more vanishing moments or motivate the search for short biorthogonal wavelets with those vanishing moments.
Bibliography


Appendix

List of Abbreviations

AV Abry-Veitch Method

$B(t)$ Brownian motion

$B_H(t)$ Fractional Brownian Motion

FBM Fractional Brownian Motion

FGN Fractional Gaussian Noise

$H$ Hurst Parameter

LRD Long-Range Dependence

SRD Short-Range Dependence

R/S Rescaled Adjusted Range

Variance In this thesis, variance and autovariance are used interchangeably

$X_t$ Long-range dependence process

$Y_t$ Self-similar process
Vita

Coming from Vietnam, Phat Piron was born in Palawan, Philippines. Almost immediately, he took off for Belgium where he spent most of his early life. In 1998, he entered the University of Liège, Belgium, to pursue an engineering degree. He received his degree of “ingénieur civil électricien, orientation informatique” in June 2003, with distinction. (This compares to a M.S. degree in Computer Engineering.) His graduating thesis was entitled “Congestion control for layered multipoint communications with multiple sources,” and was directed by Professor Guy Leduc. Next, he moved to Louisiana State University in Baton Rouge to pursue the degree of Master of Science in Electrical Engineering. He will graduate in May 2005 with a GPA of 4.0.